STRATEGIES FOR SOIL QUALITY ASSESSMENT USING VNIR HYPERSPECTRAL SPECTROSCOPY

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Masters of Science

by

Rintaro Kinoshita

January 2012

ABSTRACT

Visible and near-infrared reflectance spectroscopy (VNIRS) is a rapid and non-destructive proximal sensing method that can predict various soil properties, and has the potential to dramatically reduce the time and cost of soil analysis. In this study, the predictability of VNIRS was assessed for sixteen soil quality indicators of Western Kenyan soils. It successfully predicted SQ indicators (R^2 > 0.80; ratio of performance to deviation (RPD) > 2.00) including Ca, soil organic matter, active carbon, water content at permanent wilting point, cation exchange capacity, clay, sand, and Cu of Western Kenyan soil.

VNIRS was also employed to analyze soil organic carbon (SOC) distribution within a micro-watershed in Costa Rica. It successfully predicted SOC content with R^2 of 0.82. It appears that high spatial sampling intensities, inexpensive SOC analysis, and the "measure and multiply" extrapolation method provided by VNIRS is highly applicable for studying SOC distribution in complex agroforestry watershed.

BIOGRAPHICAL SKETCH

Rintaro Kinoshita was born in Tokyo Japan on the 8th July 1986. He lived surrounded by forests of the Boso peninsula near Tokyo in his youth where he fell in love with the trees, soils, the ocean, and farming. In 1995, he moved to Yokohama where he attended his primary and secondary school and much time was spent playing basketball. In 2002, he started attending a high school in Tokyo but later that year he decided to study abroad to immerse himself in a completely new world.

He arrived to Cardiff, Wales in August 2003 without knowing much English, and he attended a sixth form college to gain his Advanced Level General Certificate of Education. He loved Wales and it reminded him of living close to the countryside. He anticipated attending a Japanese university upon returning to Japan in July 2004, however, he decided to stay in Wales and studied at the University of Wales, Bangor for 3 years where he completed his B.Sc. in agriculture in 2008.

In April 2009, he started his post-bachelor internship with CIRAD (French Agricultural Research Institute for Development) in Costa Rica. He again enjoyed working with local farmers in the medium of Spanish and also gained valuable experience in conducting scientific research.

In January 2010, he joined the Department of Crop and Soil Sciences of Cornell University to study his master's degree.

ACKNOWLEDGEMENTS

I would like to show my profound gratitude to my major advisor Dr. Harold van Es for taking me on as a member of his research team and giving me advice on technical matters and conducting sound research. I also greatly appreciate his advice in skill development for a future career. He took a great amount of time to respond my questions despite the demand of his work as a department chair.

I am deeply grateful to Dr. Olivier Roupsard of CIRAD, France who first took me on in his research project for my post-bachelor internship in Costa Rica in 2009. He also gave me a great opportunity to conduct part of my master's research at his research site. I learnt a great deal from him as a scientist and he also kindly shared a lot of resources with me throughout.

My minor advisor Dr. Todd Walter gave me useful advice especially at the time of data presentation and throughout editing my thesis. I am also grateful to Dr. Dean Hively of USDA-ARS for his help in introducing me into the spectroscopy equipment and helping with sample analysis. I would also like to thank Dr. Volkan Bilgili of Harran University, Turkey for his support in the spectroscopy work.

I am grateful to Dr. Bianca Moebius-Clune of our research team for providing me with a great dataset of soil quality from Kenya. Without her support, I would not be able to complete the first component of this study. Also, she gave me insights into how soil quality functions within

agricultural environment and the importance of soil quality measurement for the purpose of alleviating crop production constraints.

I also thank Bob Schindelbeck for advice on research experiments and his ample experience. I always appreciated his friendliness to others and passion towards making things work.

Much appreciation goes to everyone who helped me in Costa Rica: Dr. Federico Gómez-Delgado, Alexis Perez, Dr. Tiphaine Chevallier, Dr. Glenn Galloway, Dr John Beer, Dr. Bruno Rapidel, Dr. Jean-Michel Harmand, Dr. Philippe Vaast, Dr Jacques Avelino, Dr Jeffery Jones, Dr. Francisco Jimenez, Elias De Melo, Douglas Navarro, Patricia Leandro, Cipiriano Rivera Wilson, Alfonso Robelo, Guillermo Ramirez, Rafael Acuña Vargas, Manuel Jara and to Alonso Barquero. My massive appreciation goes to Alvaro Barquerro and his family for their support in soil sampling as well as being great mental support and treating me as part of their family.

I would like to thank the Heiwa Nakajima Foundation of Japan for supporting my MS study as well as the Saltonstall Family for funding my continuing studies into a PhD program. Also, USDA-NIFA Special Grant on Computational Agriculture, the NSF Bio-Complexity Initiative (BCS-0215890), and the European project CAFNET (EuropAid/121998/C/G) for supporting part of my research.

Finally, I would like to thank my family and friends for their continuous support throughout. Without their care and love, I could not have completed this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. STRATEGIES FOR SOIL QUALITY ASSESSMENT USING VNIR HYPERSPECTRAL SPECTROSCOPY IN A WESTERN KENYA CHRONOSEQUENCE

1.1. ABSTRACT

Visible and near-infrared reflectance spectroscopy (VNIRS) is a rapid and non-destructive method that can predict various soil properties. However, its application in multidimensional soil quality (SQ) assessment in the tropics still needs to be further assessed. VNIRS (350-to-2500 nm; 1 nm resolution) was employed to analyze 227 air-dried soil samples of Ultisols from a chronosequence study in Western Kenya and assess sixteen soil quality indicators. Partial least squares regression (PLSR) was validated using the leave-ten-out method as well as independent model validation with two subsamples consisting of 70 % (n = 159) and 30 % (n = 68) of total samples. Models using raw reflectance data successfully predicted SQ indicators (R^2 > 0.80; ratio of performance to deviation (RPD) > 2.00) including Ca, soil organic matter (OM_{LOI}), active carbon (C_{act}), water content at permanent wilting point (Θ_{pwp}), cation exchange capacity (CEC), clay, sand, and Cu. Moderately-well predicted indicators ($0.50 < R^2 < 0.80$; $1.40 < RPD < 2.00$) were water stable aggregation (WSA), water content at field capacity (Θ_{fc}), pH, silt, Mg, available water capacity (AWC), penetration resistance between 15 and 45 cm (PR45) and P. Poorly predicted indicators (category C; $R^2 < 0.50$; RPD < 1.40) were S, EC, K, Zn, and penetration resistance between 0 and 15 cm (PR15). In independent model validation, only sand and Ca failed to achieve R^2> 0.80 and RPD > 2.00. Raw reflectance was generally superior over 1st derivative reflectance in the prediction of the SQ indicators, whereas the combination of raw

and 1st derivative reflectance slightly improved the prediction accuracy. Combining VNIRS with selected field- and laboratory-measured SQ indicator values increased predictability. Furthermore, VNIRS showed moderate to substantial agreement in predicting interpretive SQ scores and a composite soil quality index (CSQI). Results indicate that low cost VNIRS has good potential for substituting laboratory-based methods for many physical and biological soil quality indicators, especially when combined with inexpensively measured pH, EC, PR15, and PR45. But conventional soil chemical tests of P, K, Zn, and S may need to be retained to provide comprehensive soil quality assessments.

1.2. KEYWORDS

Soil quality; VNIR spectroscopy; PLSR; soil reflectance; Africa

1.3. INTRODUCTION

Concerns about soil degradation and related soil organic carbon losses are prevalent in Africa. The loss of soil organic carbon is related to yield loss by soil erosion, and poor nutrient and water retention, and also contributes to climate change and degradation of societal welfare (Millennium Ecosystem Assessment, 2005; Lal, 2006). In order to meet an increasing demand for food, farmers must augment agricultural output while sustainably managing their soil resources. Consequently, there is a need for local soil quality (SQ) assessment to support implementation of site-specific soil conservation measures that help alleviate SQ constraints.

Local SQ assessment is also important for monitoring sustainable agricultural land management practices to qualify for participation in carbon credit trading (The World Bank, 2010).

Doran and Parkin (1994) defined soil quality as "the capacity of a specific kind of soil to function, within natural or managed ecosystem boundaries, to sustain plant and animal productivity, maintain or enhance water and air quality, and support human quality and habitation". Assessment of SQ may be achieved by measuring SQ indicators that represent soil physical, chemical and biological properties which in turn govern soil processes (Doran and Parkin, 1994). Moebius-Clune (2010) established a framework for assessing dynamic SQ using selected indicators for smallholder agricultural soils from Western Kenya, most of which are also measured by the publicly available Cornell Soil Health Test (Idowu et al, 2008; Gugino et al. 2009). These SQ indicators include soil texture, total soil organic matter (OM_{LOI}) content, active C content (C_{act}), water stable aggregation (WSA), water content at field capacity (Θ_{fc}), water content at permanent wilting point (Θ_{pwp}), available water capacity (AWC), soil penetration resistance between 0 and 15 cm (PR15) and between 15 and 45 cm (PR45), pH, electrical conductivity (EC), cation exchange capacity (CEC), P, K, Ca, Mg, Cu, Zn and S. Interpretive SQ scores based on the measured indicators were also developed utilizing indicator- and texture-specific scoring functions to identify SQ constraints (Moebius-Clune, 2010). Having successfully measured and defined these SQ indicators and SQ constraints using field- and laboratory-based methods, we are interested in developing a more rapid, and cost effective method that requires minimal infrastructure for determining SQ.

Visible and near-infrared reflectance spectroscopy (VNIRS) was first applied in the assessment of moisture content of seeds as well as other plant products (Hart et al., 1962), and also the examination of seed and forage quality (Williams, 1975). In soil science, Chang et al. (2001) successfully predicted total organic carbon and nitrogen, gravimetric soil water content, soil water content at -1.5 MPa, exchangeable calcium, CEC, silt, and sand content with $R^2 > 0.80$. Others have also utilized VNIRS for assessment of potentially mineralizable nitrogen (Reeves and Van Kessel, 1999; Morón and Cozzolino, 2002), heavy metals, micronutrients (Kooistra et al., 2001; Cozzolino and Morón, 2003; Udelhoven et al., 2003), and AWC and WSA (Idowu et al., 2008). In addition, spatial variability within a single field can be more accurately predicted by VNIRS compared to existing methods since larger numbers of samples can be processed rapidly at low cost (Bilgili et al., 2011; Hively et al., 2011). Hand-held units for in-situ soil estimation, as well as with airborne systems are also available, which may further increase efficiency, availability, and spatial coverage (Kooistra et al., 2003; Stevens et al., 2008). VNIRS is a non-destructive technique to analyze visible (350-700 nm) and near infrared (700-2500 nm) reflectance spectra that can be interpreted using multivariate statistical or data mining techniques that relate spectra to directly-measured soil characteristics. The resulting prediction models can be used to assess soil properties of unknown samples. In the mid-infrared region, stretching and bending vibrations in molecular bonds such as C-C, C-H, N-H and O-H cause absorption of specific spectral bands of equal energy (Johnston et al., 1994; Reeves et al., 2011, in review). These molecular responses cause weak spectral overtones in the near-infrared region that can be detected through VNIRS. In the near-infrared region, physical properties of soil, such as the size and the shape of soil particles, the voids between soil particles and the arrangement of

soil particles, also affect the spectral characteristics (Reeves et al., 2011, in review). Soil properties are categorized as primary properties when their effects on spectral overtones (related to organic functional groups, particle sizes and water) are directly represented by the measured indicator (total C, total N, moisture content, particle size, and aggregation; Chang et al., 2001). Other soil properties are categorized as secondary properties, meaning that VNIRS is known to successfully predict these because they are correlated with primary properties.

Different statistical data processing methods may provide variable predictive outcomes. Some authors used the first derivatives of soil reflectance spectra (Chang et al., 2001; Reeves et al., 2002; Shepherd and Walsh, 2002; Russell, 2003; Morón and Cozzolino, 2004; Couteaux et al., 2005), while others used raw spectra (Reeves et al., 2006; Bilgili et al., 2010), or second derivatives (Salgó et al., 1998; Fystro, 2002). Diverse methods also exist for model calibration such as principal component regression (PCR), partial least squares regression (PLSR), stepwise multiple linear regression (SMLE), Fourier regression, locally weighted regression (LWR) and artificial neural networks (Creaser and Davies, 1988; Naes and Isaksson, 1990; Holst, 1992). Although Chang et al. (2001) argued that no universally perfect calibration methods exist, partial least squares regression (PLSR) is emerging as a successful method that has similar or better predictability than other methods (Viscarra Rossel et al., 2006; Bilgili et al., 2010).
In addition, Reeves et al. (2011, in review) mention that the most accurate model calibration is achieved when calibration samples are collected in close proximity to the unknowns and a wide data range is included.

The objectives of this study were to (1) examine whether VNIRS is a suitable method to rapidly assess SQ for agricultural soils from Western Kenya, (2) to determine the most suitable spectral preprocessing method for SQ indicator prediction, (3) to assess the predictability of SQ indicators by the combined uses of VNIRS and field or laboratory measured SQ indicator values, and (4) to evaluate whether single or combined uses of VNIRS is applicable to predict interpretive SQ scores and a composite soil quality index (CSQI) established as part of the Cornell Soil Health Test (Gugino et al., 2009; Moebius-Clune, 2010; Moebius-Clune et al., 2011),

1.4. MATERIALS AND METHODS

1.4.1. Location, Climate and Soil

The study site is located at the Kakamega-Nandi Forest margins in Vihiga (between 0°00' N and 0°13' N latitude and between 34°45' E and 35°03' E longitude) in Western Kenya, which is the largest remainder of Guineo-Congolean Forest in the country. The surrounding catchment feeds Lake Victoria (Lung and Schaab, 2006). Government settlement plans and poverty in the area have caused conversion of the forest into subsistence agricultural land, which is the dominant form of land use along the forest margins in the area. Annual rainfall ranges from 1800 to 2100 mm $year^{-1}$, peaking in two rainy seasons from March to August and September to January. This bimodal rainfall pattern allows for two cereal cropping periods per year.

A chronosequence study was carried out using forested sites along with surrounding continuous maize fields (Co) and kitchen gardens (Ki) that were converted from primary forest during the

1900s. The site has been used to assess changes in carbon and nutrient pools, nutrient use efficiency and SQ degradation over time after conversion to agriculture (Kimetu et al., 2008; Kinyangi, 2008; Ngoze et al., 2008; Moebius-Clune et al., 2011). The site's soils are Ultisols which contain low activity kaolinite and high contents of Fe and Al oxides (Krull et al., 2002). Two distinctive parent materials were present, biotite-gneiss in the Nandi region, and undifferentiated basement system rock, composed predominantly of Precambrian gneisses in the Kakamega region (Krull et al., 2002; Solomon et al., 2007; Kimetu et al., 2008).

The Kakamega chronosequence consists of three clustered forest sites and twelve farms converted approximately in 1930, 1950, 1970, 1985, and 1995 with elevation ranging from about 1600 to 1700 m ($\bar{x}$ = 1632 m, s = 36 m), and with coarser soil textures. The Nandi chronosequence contains three forest areas with three individual forest sites in each area and twenty-seven farms converted approximately in 1900, 1930, 1950, 1970, 1985, 1995, and 2000, with elevation ranging from 1560 to 2028 m ($\bar{x}$ = 1789 m, s = 108 m), and finer soil textures. A wide range of soil degradation dynamics over 100 years (Moebius-Clune et al., 2011) provides an ideal range in soil quality status for building a model for VNIRS analysis.

1.4.2. Management Systems

Two different long-term agricultural land management systems were sampled within each farm: Continuous maize and kitchen gardens (Moebius-Clune et al., 2011). Maize fields were tilled twice a year to a depth of 10 to 15 cm by hand hoe and received negligible amounts of fertilizer until 2004, after which N, P and K fertilizer were applied at the rate of 120, 100, 100 kg ha^{-1} per cropping season (Kimetu et al., 2008; Ngoze, 2008). Four organic matter amendment treatments

were also implemented on continuously cropped maize fields for three consecutive growing seasons starting in 2005 (Kimetu et al., 2008). Amended plots received 6t C ha^{-1} equivalent of *Tithonia diversifolia* leaves (Ti), cattle manure (Ma), wood charcoal (Ch), or sawdust (Sa) for a total of 18 t C ha^{-1} (Kimetu et al., 2008). In kitchen gardens, diverse fruit and vegetable crops were grown in polyculture receiving household organic wastes and cooking ash (Moebius-Clune et al., 2011).

1.4.3. Soil Sampling

Soil samples were collected in July and August 2007 from four primary forest areas, and from farm fields of all conversion years and six treatments of Ki, Co, Ti, Ma, Ch, Sa, from the 0-to-15 cm depth using a Dutch type soil auger after the removal of surface residue (Moebius-Clune, 2010). At each sampling point, five sub samples were collected and then mixed to obtain a representative sample. Collected soil samples were air-dried and passed through a 2-mm sieve.

1.4.4. Soil Quality Assessment

Soil quality was assessed by Moebius-Clune (2010) using a modified set of indicators from the Cornell Soil Health Test (Idowu et al., 2008; Gugino et al., 2009). Soil penetration resistance was measured in the field using a soil compaction tester (DICKEY-john Corporation; Auburn, IL), for surface (0-to-15 cm, PR15) and subsurface soils (15-to-45 cm, PR45). The remaining indicators were measured using laboratory procedures. Soil texture, an inherent soil property that adjusts the interpretation of dynamic SQ, was assessed using a rapid quantitative method

developed by Kettler et al., (2001). The soil sample was dispersed with 3% hexametaphosphate ($(NaPO_3)_n$) and a combination of sieving and sedimentation steps was used to separate size fractions. WSA was assessed using a rainfall simulator (Ogden et al., 1997) that allows the soil particles to slake under known rainfall energy, applying 2.5 J of energy for 300 s on aggregates (0.25 - 2 mm) placed on a 0.25 mm mesh sieve 21. The fraction of soil aggregates remaining on the sieve, corrected for stones >0.25 mm, was regarded as the percent WSA after drying at 105 °C (Gugino et al., 2009). Θ_{fc}, Θ_{pwp}, and AWC were assessed gravimetrically. Saturated soil samples were equilibrated to pressures of 10 kPa (Θ_{fc}) and 1500 kPa (Θ_{pwp}) on two ceramic high pressure plates (Topp et al., 1993). The difference in moisture content between these two pressure points was considered as the AWC. OM_{LOI} was analyzed by loss on ignition in a muffle furnace (350°C for 18 hours). The low temperature prevents structural water loss, which is known to occur at 450 – 600°C (Ball, 1964; Rhodes et al., 1981). C_{act} was measured by oxidizing samples with dilute potassium permanganate ($KMnO_4$) followed by measuring absorbance at 550 nm using a hand-held colorimeter (Hach, Loveland, CO). C_{act} is recognized to be very sensitive to soil management and correlated with soil microbial activity, aggregation and crop yield (Islam and Weil, 2000; Weil et al., 2003; Mtambanengwe et al., 2006). A 1: 2.5 soil and water suspension was assessed for pH and EC using a hand-held portable probe (SM802 Smart Combined Meter, Milwaukee Industries, Inc., Rocky Mount, NC). Soil pH and EC are overall indicators of nutrient availability, and EC has also been used as an indicator of soil nitrate concentration and soil salinity (Arnold et al., 2005; Bastida et al., 2008; Wongpokhom et al., 2008). Soil nutrients, including P, K, Mg, Ca, Zn, Cu and S, were extracted using Mehlich-3 (Mehlich, 1984) and quantified by plasma optical emission spectrometry (ICP-OES, Varian 730-

ES, Mulgrave, Victoria, Australia), and CEC was calculated as the sum of K, Mg, Ca and exchangeable acidity by the Agricultural Analytical Services Laboratory (Pennsylvania State University, University Park, PA).

1.4.5. Data Interpretation and Scoring Curves

Measured values of SQ indicators can be interpreted using non-linear scoring functions to assess the degree to which soil processes may be constrained. The development of the scoring curves used for this study was described by Moebius-Clune (2010). Briefly, three types of scoring functions were constructed: 'less is better' (for PR15 and PR45), 'optimum range' (for pH) and 'more is better' (for OM_{LOI}, C_{act}, WSA, AWC, EC, P, and K; Karlen et al., 1994). A cumulative normal distribution function was applied to each measured SQ indicator to calculate SQ scores:

$$CND(x, m, s) = 100\frac{1}{2}\left(1 + \text{erf}\left[\frac{(x-m)}{s\sqrt{2}}\right]\right) \qquad [1]$$

where x is the SQ indicator value, m is the population mean, s is the standard deviation of the population, and erf denotes the error function. Clay content (<15% or >15%) was utilized to select a texture-based scoring function for WSA, AWC, PR45, OM_{LOI}, C_{act}, and K since the interpretation of these measured values must take into account inherent effects of soil particle size (Dexter, 2004; Moebius et al., 2007).

SQ scores for each indicator were interpreted to signify optimum ($70 \leq$ score ≤ 100), intermediate ($30 <$ score < 70) and constrained ($0 \leq$ score ≤ 30) functioning of soil processes. A composite soil quality index (CSQI) was calculated as the mean value of the above ten individual

SQ scores, and provides an overall assessment of soil quality. The CSQI values were qualitatively interpreted as CSQI < 40 (very low), 40 < CSQI < 55 (low), 55 < CSQI < 70 (medium), 70 < CSQI < 85 (high), and CSQI > 85 (very high).

1.4.6. Visible and Near-Infrared Reflectance Spectroscopy (VNIRS)

The reflectance of soil samples was determined in both the visible and near infrared spectral regions between 350 nm and 2500 nm at 1 nm resolution using a Fieldspec Pro hyperspectral sensor (Analytical Spectral Devices, INC., Boulder, CO). Air-dried samples were placed in a 4 cm diameter optical quality petri dish, and spectral reflectance was collected through the glass bottom at a constant angle (55 degrees from horizontal) from a distance of 4 cm, in an enclosed box. A tungsten quartz halogen lamp was used as a light source (Muglight sensor attachment; Analytical Spectral Devices, INC., Boulder, CO), and the reflectance was measured using a fiberoptic cable. After 50 readings, the sample petri dish was rotated 90 degrees and another 50 readings were taken, which were averaged to produce two spectra per sample. The signal was optimized before the measurement and the accuracy of the detector was checked every 8 samples using a white Spectralon standard. When the white Spectralon reading was not stable, the instrument was recalibrated.

1.4.7. Spectrum Preprocessing

Reflectance data were exported from binary to ASCII file format using ViewSpec Pro software (Analytical Spectral Devices, INC., Boulder, CO). The two spectral readings per sample

obtained during measurement were averaged using Unscrambler 10.0.1 (CAMO software, Oslo, Norway, 2010). Derivative values were used to overcome baseline noise and to amplify spectral signals (Reeves et al., 2002). First derivatives were calculated using the Savitsky-Golay transformation (Savitzky and Golay, 1964), which performs a second order polynomial regression using 4 points to the left and one point to the right side of each point for smoothing. This transformation method is suitable for VNIRS applications since it removes noise from the data while retaining original spectral characteristics with minimum distortion (Ruffin and King, 1999). Wavelengths with low signal to noise ratio (350-420 nm, 961-1019 nm, 1771-1829 nm, and 2481-2500), resulting from splicing between individual spectrometers were omitted before data analysis.

1.4.8. VNIRS Modeling

The VNIRS analysis utilizes empirical models based on correlations between soil reflectance and soil property values. Calibration models for predicting SQ indicator values were constructed utilizing the partial least squares regression (PLSR) method, which is commonly applied in VNIRS studies for predicting wide ranges of soil properties (Bilgili et al., 2010; Kusumo et al., 2010; Hively et al., 2011). It relates two variables, X (spectral readings) and Y (measured soil property values), by a linear multivariate model. Orthogonal and weighted linear combinations of the spectral readings are used for predicting each Y variable. PLSR is suited to handle data with strong collinearity in the (X) variables, which are usually more numerous than the observations (Y) that they predict. The selection of the number of factors (F) to include was critical to avoid over-fitting or under-fitting of the model to the data. Over-fitting of calibration models reduces

their ability to predict soil properties of new unknown soil samples. In this study the smallest number of F that produces sufficiently small root mean square error of prediction (RMSEP) was selected by analyzing the total residual variance plot. Also, the maximum F was restricted to 20. For the purpose of determining the most suitable spectral preprocessing method, the leave-ten-out cross validation method was used (Heise et al., 2002). Separate validation sample subsets (Brunet et al., 2007; Bilgili et al., 2010) were also used to assess the predictive ability of each model, using the best preprocessing method. For this purpose, samples were divided randomly into two subsets consisting of 70 % (validation set 1, n = 159) and 30 % (validation set 2, n = 68) of the total samples and each of the subsets was used once for each calibration and validation.

1.4.9. Prediction Accuracy

The accuracy of each PLSR model for SQ indicator prediction was evaluated using the coefficient of determination (R^2) of measured vs. VNIRS-predicted values. RMSEP, a measure of accuracy and precision calculated as the differences between model-predicted values and observed values, was determined as shown in Eq. [2];

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(Y_{pred.} - Y_{meas.})^2}{n-1}} \qquad [2]$$

The ratio of performance to deviation (RPD) was also calculated, which is the ratio of standard deviation to standard error of prediction in each sample. It is a measure of the ability of a VNIRS-based model to predict each SQ indicator (Williams and Sobering, 1993).

1.4.10. Combined Use in Modeling of VNIRS and Selected Soil Properties

Research has generally found poor predictability by VNIRS of soil chemical properties such as K, Zn, Na, and EC (Chang et al., 2001; Awiti et al., 2008; Idowu et al., 2008; Bilgili et al., 2010), which constrains its application as a tool for soil nutrient analysis. Therefore, we evaluated the combined uses of VNIRS with selected soil nutrient laboratory measurements to predict other SQ indicators through the PLSR model. Morgan et al. (2009) combined the VNIRS values with soil color, soil pH, and soil reaction to 1M HCl to predict organic and inorganic C, which resulted in higher predictability compared to VNIRS alone. In this study, the selected auxiliary measurements were soil penetration resistance (PR15 and PR45), soil pH, and EC, which require only minimal investments in equipment and can be operated in the field. In addition, conventional soil nutrient analyses may be combined with VNIRS where a conventional soil-testing laboratory is well established, which sometimes is the case in Africa.

The tested combinations of predictors were 1) raw reflectance VNIRS, plus PR15, PR45, pH, and EC, and 2) same as 1), plus soil chemical indicator values (CEC, P, K, Ca, Mg, Cu, Zn, and S).

1.4.11. Categorical Prediction

Understanding soil process constraints for each soil sample is important as this can guide development of site-specific soil management strategies (Moebius-Clune, 2010). Having classified each SQ score into three categories (optimum, intermediate and constrained) or five categories for the CSQI (very high, high, medium, low and very low), it was assessed whether raw reflectance VNIRS-PLSR models are capable of predicting these classifications with

sufficient accuracy. The VNIRS-predicted values were used to calculate specific SQ scores through the non-linear scoring functions presented by Moebius-Clune (2010). In addition, the best combined VNIRS-laboratory prediction model was compared to the raw reflectance VNIRS-PLSR model for SQ score analysis. The kappa statistic (κ), which measures the percentage agreement between two different analysis methods by testing the hypothesis that the agreement between the two methods occurred just by chance (Viera and Garrett, 2005), was applied for this purpose. The kappa statistic, which has been used to validate the quality of various remote sensing-based predictions against direct measurement (Aspinall, 2002; Chikhaoui, et al., 2005; Bilgili et al., 2010), is determined as;

$$\kappa = \frac{(p_o - p_e)}{(1 - p_e)} \qquad [3]$$

where p_o is percentage observed agreement and p_e is expected agreement by chance alone (Viera and Garrett, 2005). A κ value of 1 indicates perfect agreement whereas a κ of 0 refers to agreement just by chance. The percentages of correctly classified samples, as well as the kappa statistic were calculated using Microsoft Excel (Microsoft Corporation, Redmond, WA, 2010).

1.5. RESULTS AND DISCUSSION

1.5.1. Soil Properties

Summary statistics of the SQ indicators show a wide range of values for each indicator due to the nature of the chronosequence study design, which represents 100 years of soil degradation dynamics (Table 1.1). The wide range of values used as an input for PLSR was favorable for

building a rigorous model as discussed by Reeves et al. (2011, in review). Ca had the highest concentration of all the exchangeable cations (mean = 1675 mg kg^{-1}) and therefore occupied the highest fraction of the CEC. In our study, several indicators were hypothesized to be primary properties (i.e. properties assumed to be directly measured by VNIRS via reflectance from soil physical make-up or molecular bonds) such as sand, silt clay, OM_{LOI}, C_{act}, and WSA and the remaining measured indicators as secondary properties (Table 1.2; Chang et al., 2001). Pearson correlation coefficients indicated significant correlations ($\alpha = 0.01$ or $\alpha = 0.05$) between some secondary properties and primary properties.

In our study, Θ_{fc}, $\Theta_{pwp,}$ and AWC were strongly correlated to texture and C_{act} since water is stored in soil pores and active soil organic matter, but less correlated to $OM_{LOI,}$ which is an indicator of total organic matter content. CEC and Ca were also strongly correlated with C_{act} ($r = 0.86$ and 0.87) as a result of high CEC in the soil organic matter fraction itself. CEC was also strongly correlated with Ca and Mg ($r = 0.83$), in part owing to the fact that it was measured as the sum of K, Mg and Ca. These significant correlations can partly explain the acceptable predictability of VNIRS for secondary properties (Ben-Dor and Banin, 1995). Bilgili et al. (2010) similarly showed high correlations of clay and CEC to Ca and Mg content similarly, and suggested that the good predictability of secondary properties can be explained by these correlations.

1.5.2. Soil Reflectance

Raw reflectance spectra (Fig. 1.1a) and the first derivatives (Fig. 1.1b) are shown for all 227 soil samples. As is typical of soils, the reflectance was lower in the visible range and increased in the near-infrared range between 700 nm and 1800 nm. There were three distinctive absorbance

Table 1.1 Descriptive statistics of the measured soil quality indicators (n=227).

	Min	Max	Mean	St.dev
Sand (%)	8	72	44	15
Silt (%)	16	64	38	10
Clay (%)	4	45	18	9
OM_{LOI}[a] (g kg^{-1})	24.9	208.7	77.3	37.7
C_{act}[b] (mg kg^{-1})	24	1306	483	283
WSA[c] (%)	21	99	61	20
Θ_{fc}	0.18	.52	0.32	0.07
Θ_{pwp}	0.07	0.35	0.17	0.05
AWC[d] (m^3 m^{-3})	0.07	0.26	0.15	0.04
PR15[e] (kPa)	9	2172	207	227
PR45[f] (k a)	393	3103	1404	668
pH	4.90	7.95	6.04	0.59
EC[g] (dS m^{-1})	0	0.31	0.04	0.05
CEC (cmol kg^{-1})	7	28	16	5
P (mg kg^{-1})	4	205	37	30
K (mg kg^{-1})	2	1250	287	184
Ca (mg kg^{-1})	150	5016	1675	1074
Mg (mg kg^{-1})	28	664	219	133
Cu (mg kg^{-1})	0.71	7.69	2.85	1.51
Zn (mg kg^{-1})	1.48	78.22	11.6	11.79
S (mg kg^{-1})	9.81	28.09	14.52	2.90

[a] OM_{LOI} = total organic matter by loss on ignition at 350 C.

[b] C_{act} = permanganate-oxidizable, biologically active carbon.

[c] WSA = water stable aggregation.

[d] AWC = available water capacity

[e] PR15 = penetration resistance between 0 and 15 cm.

[f] PR45 = penetration resistance between 15 and 45 cm.

[g] EC = electrical conductivity

Table 1.2 Pearson correlation coefficients for the measured soil quality indicators

	Sand[a]	Silt[a]	Clay[a]	OM_{LOI}[a]	Cact[a]	WSA[a]	Θ_{fc}[b]	Θ_{pwp}[b]	AWC[b]	PR15[b]	PR45[b]	pH[b]	EC[b]	CEC[b]	P[b]	K[b]	Ca[b]	Mg[b]	Cu[b]	Zn[b]	S[b]
Sand[a]	1.00																				
Silt[a]	-0.80**	1.00																			
Clay[a]	-0.79**	0.26**	1.00																		
OM_{LOI}[a]	ns	0.29**	-0.26**	1.00																	
Cact[a]	ns	0.18**	-0.31**	0.23**	1.00																
WSA[a]	-0.21**	0.15*	0.18**	ns	0.67**	1.00															
Θ_{fc}[b]	-0.50**	0.63**	0.15*	0.26**	0.71**	0.55**	1.00														
Θ_{pwp}[b]	-0.62**	0.49**	0.50**	ns	0.55**	0.76**	0.80**	1.00													
AWC[b]	ns	0.44**	-0.36**	0.36**	0.52**	ns	0.69**	ns	1.00												
PR15[b]	ns	ns	0.17**	ns	ns	0.14*	ns	ns	-0.29**	1.00											
PR45[b]	ns	ns	0.19**	ns	-0.35**	-0.16*	-0.20**	ns	-0.23**	0.54**	1.00										
pH[b]	ns	0.20**	-0.27**	0.15*	0.53**	0.23**	0.39**	0.18**	0.43**	ns	ns	1.00									
EC[b]	0.17**	ns	-0.33**	0.16*	0.61**	0.40**	0.43**	0.29**	0.35**	ns	ns	0.71**	1.00								
CEC[b]	ns	0.23**	ns	0.16*	0.86**	0.67**	0.71**	0.65**	0.39**	ns	-0.30**	0.45**	0.45**	1.00							
P[b]	0.41**	-0.24**	-0.41**	ns	ns	ns	-0.29**	-0.39**	ns	0.21**	0.34**	0.31**	0.41**	-0.14*	1.00						
K[b]	ns	ns	ns	ns	0.47**	0.50**	0.33**	0.42**	ns	ns	ns	0.58**	0.63**	0.52**	0.27**	1.00					
Ca[b]	ns	0.17*	-0.27**	0.16*	0.87**	0.51**	0.66**	0.47**	0.53**	ns	-0.29**	0.76**	0.65**	0.83**	ns	0.50**	1.00				
Mg[b]	ns	0.28**	ns	0.19**	0.79**	0.58**	0.73**	0.61**	0.47**	ns	-0.15*	0.71**	0.71**	0.83**	ns	0.62**	0.86**	1.00			
Cu[b]	ns	0.27**	-0.26**	0.20**	-0.35**	-0.63**	ns	-0.40**	0.33**	ns	ns	ns	ns	-0.49**	ns	-0.36**	-0.23**	-0.21**	1.00		
Zn[b]	ns	ns	-0.21**	0.28**	0.34**	0.18**	0.20**	ns	0.21**	ns	ns	0.59**	0.62**	0.27**	0.43**	0.44**	0.40**	0.44**	ns	1.00	
S[b]	ns	ns	ns	ns	0.15*	0.21**	0.16*	0.19**	ns	0.14*	0.21**	ns	0.45**	ns	0.26**	ns	ns	ns	ns	0.17*	1.00

**,* Significant at $\alpha = 0.01$ and 0.05, respectively

ns = not statistically significant at $\alpha = 0.05$

[a] = primary property

[b] = secondary property

bands at 1400, 1900 and 2200 nm. The spectral band at 1400 nm is strongly affected by OH features of free water and clay lattices whereas the band at 1900 nm is affected by OH features of free water and 2200 nm by OH features of clay lattices (Hunt, 1980).

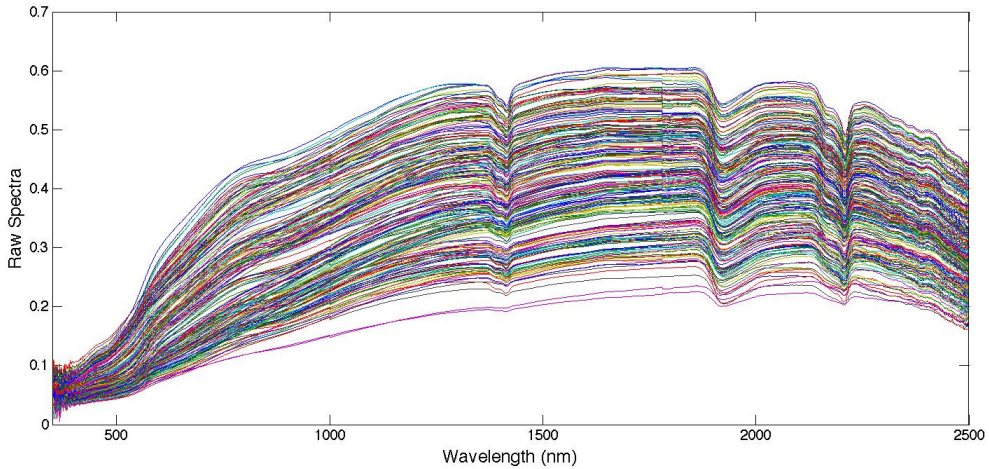
Albedos for the entire sample set showed a maximum value of 60% (Fig. 1.1a), which is comparable to a study on similar Kenyan soils (Nitisols; Awiti et al., 2008). Generally, higher soil organic matter contents, and thus darker soils, appear lower on the raw spectral reflectance axis (Stoner and Baumgardner, 1981; Ben-Dor et al., 1999; Vagen et al., 2006). Particle sizes alter the albedo as well, with larger particles showing lower reflectance of incoming light, although this relationship is partly affected by soil mineralogy (Bowers and Hanks, 1965; Viscarra Rossel and McBratney, 1998a,b). Our spectral data had a wide range of reflectance due to significant variability in SQ representing the dynamics of soil degradation (Fig. 1.1a)

1.5.3. Prediction of Soil Properties

The leave-ten-out cross-validation method was applied to compare three spectral pre-processing methods for predicting SQ indicators. Prediction accuracy was assessed by R^2, RMSEP and RPD (Table 1.3). RPD was adopted in our study, since RMSEP alone cannot provide sufficient information on model predictability due to the variable SD of the SQ indicators, while RPD can be compared across properties measured in different units. No critical value exists for RPD in soil science, but an RPD > 2 is denoted as satisfactory (Chang et al., 2001). We adopted three categories of predictability suggested by Chang et al. (2001), which were category A ($R^2 > 0.80$; RPD > 2.00), category B ($0.50 < R^2 < 0.80$; $1.40 <$ RPD < 2.00), and category C ($R^2 < 0.50$; RPD

Fig. 1.1 Soil raw spectra (a) and first derivative spectra (b) for the full data set (n=227)

(a)



(b)

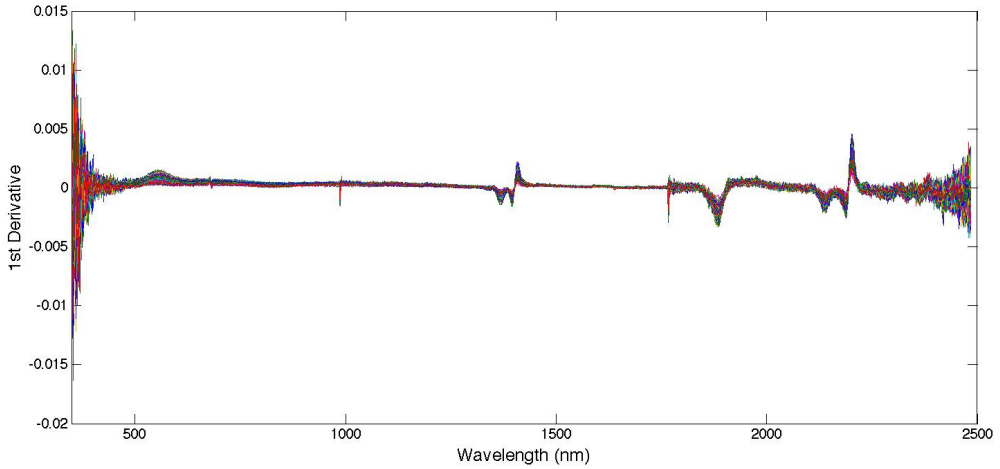


Table 1.3 Statistical results of cross-validation of indicator prediction based on VNIRS[a] with PLSR[b] using raw reflectance, first derivative or combination of both.

	Raw Spectra				First Derivative Spectra				Raw + First			
	F[c]	R^{2d}	RMSEP[e]	RPD[f]	F	R^2	RMSEP	RPD	F	R^2	RMSEP	RPD
Sand	14	0.84	6.04	2.50	6	0.74	7.69	1.97	14	0.80	6.11	2.48
Silt	14	0.73	5.03	1.92	7	0.59	6.21	1.56	15	0.74	4.95	1.95
Clay	10	0.87	3.35	2.80	4	0.85	3.61	2.59	10	0.87	3.38	2.77
OM_{LOI}	10	0.91	11.26	3.35	6	0.88	13.06	2.89	10	0.91	11.08	3.40
C_{act}	10	0.89	92.44	3.07	5	0.84	112.45	2.52	10	0.90	90.45	3.13
WSA	8	0.78	9.18	2.15	6	0.80	8.90	2.22	8	0.79	9.17	2.15
Θ_{fc}	11	0.77	0.04	2.09	7	0.71	0.04	1.85	11	0.77	0.03	2.10
Θ_{pwp}	14	0.89	0.02	2.98	7	0.83	0.02	2.41	13	0.88	0.02	2.90
AWC	n/a	0.63	0.03	1.62	n/a	0.61	0.03	1.59	n/a	0.60	0.03	1.53
PR15	8	0.15	210.53	1.08	2	0.12	214.32	1.06	8	0.14	210.93	1.08
PR45	17	0.62	411.78	1.62	5	0.51	471.78	1.42	17	0.62	410.27	1.63
pH	19	0.76	0.29	2.04	6	0.54	0.40	1.47	19	0.77	0.29	2.05
EC	10	0.44	0.03	1.34	3	0.29	0.04	1.18	10	0.44	0.03	1.34
CEC	10	0.88	1.82	2.88	3	0.84	2.11	2.47	10	0.88	1.81	2.89
P	18	0.59	19.53	1.55	5	0.28	25.66	1.18	18	0.60	19.19	1.58
K	4	0.31	153.49	1.20	2	0.26	159.12	1.16	4	0.30	154.19	1.19
Ca	15	0.91	328.57	3.27	6	0.83	440.79	2.44	15	0.90	340.77	3.15
Mg	10	0.72	70.12	1.90	4	0.63	81.01	1.65	10	0.72	70.72	1.89
Cu	[f]10	0.83	0.61	2.47	6	0.83	0.62	2.45	10	0.84	0.60	2.50
Zn	5	0.22	10.45	1.13	2	0.11	11.19	1.05	4	0.20	10.53	1.12
S	13	0.45	2.14	1.35	3	0.30	2.45	1.19	13	0.47	2.11	1.38

[a] VNIRS = visible and near infrared spectroscopy

[b] PLSR = partial least squares regression

[c] F = number of factors used in PLSR

[d] R^2 = coefficient of correlation

[e] RMSEP = root mean square of prediction

[f] RPD = ratio of performance to deviation

< 1.40). Whenever either one of the criteria failed to comply with a category, the indicator was placed in a lower category. Overall, best predictions were achieved with raw reflectance spectra for nine out of nineteen indicators, which categorized Ca, OM_{LOI}, C_{act}, CEC, Θ_{pwp}, clay, sand, and Cu in category A, WSA, Θ_{fc}, pH, silt, Mg, AWC, PR45, and P in category B, and S, EC, K, Zn, and PR15 in category C. The indicators in category A are either primary properties or secondary properties having significant Pearson correlations (Table 1.2) with the primary properties. The indicators in category B may be better predicted by combining VNIRS and predictive conventional analytical methods as discussed below whereas the category C indicators will not be accurately predicted by VNIRS. The poor prediction performance in category C could be caused by the result of poor correlations to primary properties as discussed previously (Table 1.2). The predictabilities in category C were in line with the study by Chang et al. (2001) on Zn (R^2 = 0.44) and by Bilgili et al. (2010) on EC (R^2 = 0.27) and K (R^2 = 0.38), where the authors also cite poor correlations between the predicted soil properties and the primary properties.

Using 1st derivative reflectance, the models for predicting Ca, OM_{LOI}, C_{act}, CEC, Θ_{pwp}, clay, Cu, and WSA remained in category A but a better predictability was only found for WSA compared to raw reflectance (Table 1.3). This is in agreement with Bilgili et al. (2010), who showed better predictability with raw reflectance for soil organic matter and soil texture using the same experimental set up. As an example, OM_{LOI} content was well predicted by the raw reflectance method (R^2 = 0.91; RPD = 3.35) but the predictability declined slightly using the 1st derivative method (R^2 = 0.88; RPD = 2.89). The need for deriving 1st derivative reflectance was minimal for this experimental setup, since the illumination was kept constant at all times, and an average

of 100 spectral readings from two different positions were taken. The combined raw + 1st derivative method had slightly higher RPD in OM_{LOI} (3.40), C_{act} (3.13), CEC (2.89), Cu (2.50), pH (2.05), PR45 (1.63), silt (1.95), P (1.58), and S (1.38). Nevertheless, improvements in prediction were generally small, and we conclude that the simple raw reflectance data are the most suitable for prediction.

The RMSEP of OM_{LOI} was relatively large (11.26 g kg^{-1}) in our study compared to 2.90 g kg^{-1} by Bilgili et al. (2010) and 7.80 g kg^{-1} by Lamsal et al. (2009), which may be related to the large SD in our data. In this study, roughly 900 km^2 of land was covered consisting of two distinctive parent materials. Past research has shown some difficulties in gaining satisfactory statistical predictions when including different parent materials (Reeves and van Kessel, 1999; Shepherd and Walsh, 2002). In order to alleviate site variability, the samples were divided according to their parent materials, however, this did not improve the RMSEP (data not shown). Shepherd and Walsh (2002) also hypothesized that high values in soil properties generally present higher RMSEP that may be caused by error in the laboratory analytical methods instead of poor prediction power.

1.5.4. Independent Sample Model Validation

The cross-validation method tends to provide overly-optimistic assessments of model performance (Dardanne et al., 2000). Therefore, independent sample sets were used for validation analyses. Model performances for each SQ indicator are shown in Table 1.4. When 70 % of the total samples were used for calibration, the R^2 values increased for OM_{LOI}, C_{act},

WSA, Mg, Cu, and S compared to the cross-validation models (validation set 1). However, the R^2 values were lower compared to the cross-validation models in all of the indicator values when only 30 % of the total samples were used for calibration (validation set 2). Indicator predictability decreased when fewer samples were assigned to the model's calibration because when the sample size is small (<100), outliers can easily affect the model (Bishop and McBratney, 2001). The lower predictability in validation set 2, therefore could be the result of spectral or laboratory outliers, which would be less significant in validation set 1. For the latter, OM_{LOI}, C_{act}, Θ_{pwp}, CEC, Cu, WSA, Ca, and clay were placed in category A. The threshold of RPD > 2 was maintained for all category A indicators and several additional indicators, including sand and Mg.

1.5.5. Combined Use of VNIRS and Conventional SQ Analyses

The predictabilities of soil chemical properties (pH, EC, P, K, Mg, Zn and S) as well as some soil physical properties (silt, WSA, AWC, Θ_{fc}, PR15 and PR45) were low using VNIRS alone and assigned in category B or C. Therefore, we evaluated the combined use of raw reflectance VNIRS with selected field or laboratory measured SQ indicator values. We tested two combinations of factors in PLSR models: 1) raw reflectance VNIRS combined with pH, EC, PR15, and PR45, and 2) and the same as 1) plus the soil chemical indicator values of CEC, P, K, Ca, Mg, Cu, Zn, and S. The RPD values were improved compared to VNIRS alone for C_{act} (3.61 vs. 3.07), sand (2.56 vs. 2.50), WSA (2.25 vs. 2.15), Mg (2.21 vs. 1.90), silt (1.95 vs. 1.92), S (1.82 vs. 1.35), K (1.54 vs. 1.20), and Zn (1.31 vs. 1.13) when measuring pH, EC, PR15, and PR45 as auxiliary

Table 1.4 Statistical results of independent sample validation of VNIRS prediction of indicators with PLSR using raw reflectance.

	Validation set 1 (ncal = 159, nval = 68)						Validation set 2 (ncal = 68, nval = 159)					
	F	R^2 cal.	$RMSEP_{cal}$[a]	R^2 val.	$RMSEP_{val}$[b]	RPD	F	R^2 cal.	$RMSEP_{cal}$	R^2 val.	$RMSEP_{val}$	RPD
Sand	14	0.84	6.33	0.79	6.03	2.22	10	0.57	8.75	0.75	7.94	1.99
Silt	15	0.74	5.14	0.66	4.97	1.72	9	0.37	6.85	0.53	6.88	1.46
Clay	8	0.86	3.67	0.80	3.79	2.26	6	0.77	4.15	0.83	4.01	2.42
OM_{LOI}	10	0.89	12.42	0.94	9.49	4.27	10	0.93	10.55	0.90	11.80	3.10
C_{act}	10	0.89	94.09	0.90	90.69	3.24	10	0.90	93.78	0.87	99.28	2.82
WSA	8	0.76	9.84	0.83	7.99	2.43	5	0.77	9.46	0.70	10.93	1.82
Θ_{fc}	10	0.74	0.04	0.77	0.04	2.12	8	0.68	0.04	0.65	0.04	1.70
Θ_{pwp}	14	0.88	0.02	0.89	0.02	3.02	8	0.79	0.02	0.79	0.02	2.19
AWC	n/a	0.55	0.03	0.19	0.04	0.95	n/a	0.21	0.04	0.53	0.03	1.41
PR15	15	0.14	241.64	n/a	140.23	0.81	15	0.23	100.85	0.02	255.47	1.02
PR45	17	0.56	465.98	0.51	395.13	1.43	14	0.62	352.09	0.41	538.05	1.31
pH	16	0.70	0.32	0.66	0.34	1.73	3	0.42	0.45	0.41	0.45	1.30
EC	10	0.44	0.03	0.40	0.04	1.30	16	0.34	0.04	0.26	0.04	1.16
CEC	8	0.85	2.04	0.86	1.98	2.65	4	0.82	2.28	0.78	2.42	2.16
P	18	0.56	18.70	0.54	23.28	1.48	4	0.19	31.40	0.05	27.36	1.03
K	2	0.31	165.23	0.02	145.43	1.02	5	0.25	128.68	0.18	178.77	1.11
Ca	10	0.85	421.22	0.82	442.81	2.39	4	0.76	525.88	0.71	585.85	1.85
Mg	12	0.72	69.83	0.76	68.27	2.07	5	0.63	86.63	0.60	82.21	1.59
Cu	10	0.82	0.69	0.85	0.48	2.64	10	0.80	0.57	0.82	0.68	2.35
Zn	4	0.20	9.46	0.21	12.61	1.13	3	0.08	13.83	0.13	9.84	1.07
S	10	0.34	2.43	0.55	1.84	1.50	11	0.66	1.61	0.23	2.60	1.14

[a] $RMSEP_{cal}$ =Root mean square error of prediction by cross-validation

[b] $RMSEP_{val}$[b=] Root mean square error of prediction by validation using independent sample set

variables (Table 1.5). The addition of soil chemical indicator values that are measured in standard soil nutrient tests further increased the RPD value of WSA (2.29 vs. 2.25), Θ_{fc} (2.44 vs. 2.09), and AWC (1.74 vs. 1.62). Improved predictabilities are in part due to either significant positive or negative Pearson correlations between the improved SQ indicators and the combined predictor variables (Table 1.2). The incorporation of the direct measurements of pH, EC, PR15 and PR45 is practical since they are not well predicted by VNIRS and also can be measured rapidly, at low cost, and even in the field. The increase in predictability moved Mg from category B to A and S and K from category C to B. Although the incorporation of soil chemical indicators is favorable, the access to well-established laboratories is limited for smallholder farmers in sub-Saharan Africa. But if traditional chemical soil testing is available, its combined use with VNIRS allows for effective expansion of the suite of predictable SQ indicators for physical and biological processes.

1.5.6. Categorical Analysis

For the purpose of diagnosing SQ constraints or landscape-wide SQ status, standardized SQ classes are perhaps more practical than numerical results of quantitative soil property values. Each of the measured SQ indicator values were scored and classified into three soil quality classes (optimum, intermediate and constrained; Moebius-Clune, 2010) using 1) existing SQ analysis methods, 2) raw reflectance VNIRS-PLSR models, and 3) best prediction models including different spectral pre-processing methods (raw, first and raw +first) and combined use of VNIRS and measured SQ indicator values (Table 1.6). The agreement of the predicted classes by methods 2) and 3) were compared to method 1) using kappa statistics. A commonly used

Table 1.5 Statistical results of cross-validation of the combined uses of VNIRS and pH, EC, PR15, PR45, and chemical indicators as predictors with PLSR.

	VNIRS + pH, EC, PR15, PR45				**VNIRS+pH,EC,PR15,PR45, nutrient contents**			
	F	R2	RMSEP	RPD	F	R2	RMSEP	RPD
Sand	18	0.85	5.91	2.56	20	0.82	6.44	2.35
Silt	18	0.74	4.95	1.95	20	0.70	5.30	1.82
Clay	9	0.84	3.73	2.51	18	0.87	3.43	2.73
OM_{LOI}	12	0.91	11.58	3.26	8	0.81	16.65	2.26
C_{act}	14	0.92	78.60	3.61	8	0.85	108.96	2.60
WSA	12	0.80	8.78	2.25	18	0.81	8.64	2.29
Θ_{fc}	14	0.77	0.03	2.09	18	0.83	0.03	2.44
Θ_{pwp}	17	0.88	0.02	2.92	13	0.87	0.02	2.78
AWC	n/a	0.62	0.03	1.61	n/a	0.67	0.02	1.74
PR15	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
PR45	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
pH	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
EC	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
CEC	9	0.83	2.15	2.43	n/a	n/a	n/a	n/a
P	15	0.58	19.76	1.53	n/a	n/a	n/a	n/a
K	10	0.58	119.32	1.54	n/a	n/a	n/a	n/a
Ca	8	0.89	356.03	3.02	n/a	n/a	n/a	n/a
Mg	8	0.80	60.39	2.21	n/a	n/a	n/a	n/a
Cu	11	0.84	0.61	2.46	n/a	n/a	n/a	n/a
Zn	8	0.41	9.01	1.31	n/a	n/a	n/a	n/a
S	14	0.70	1.59	1.82	n/a	n/a	n/a	n/a

scale for categorizing kappa (Table 1.7; Viera and Garrett, 2005) was used as a guideline. Soil texture played a role in selecting the appropriate scoring function for most SQ indicators. Therefore, the agreement for clay content between the two methods (laboratory analytical method vs. raw reflectance VNIRS) was first assessed. The two classes of clay contents (< 15% and > 15%) were assessed for kappa statistics (Table 1.8), and showed substantial agreement between the two methods. Using VNIRS-predicted values and clay content classes, OM_{LOI} and C_{act} were in substantial agreement ($\kappa > 0.61$); pH and EC in moderate agreement ($\kappa > 0.41$); WSA, AWC, P, and CSQI in fair agreement ($\kappa > 0.21$), K and PR45 in slight agreement ($\kappa > 0.01$), and PR15 is in less than chance agreement ($\kappa < 0$; Table 1.8). The best-combined prediction models improved the agreement on C_{act}, OM_{LOI}, AWC, and CSQI, while PR15, PR45, pH, EC, P and K results were obtained from traditional methods. In particular, C_{act}, AWC, and CSQI showed meaningful improvements compared to raw reflectance VNIRS-only (Table 1.8). Nevertheless, the feasibility of using some of the best-combined models is dependent on the availability of analytical resources. Lal (2006) suggests that the loss of soil organic matter is strongly linked to yield loss, especially in Africa, and is possibly the most important SQ constraint. In our study, C_{act}, which is a fraction of soil organic matter, was an important predictor of grain yield, explaining 16 and 13% of the variability in 2005 and 2007, respectively (Moebius-Clune, 2010). C_{act} represents soil biological activity that releases organic-matter-derived nutrients, which is especially valuable in smallholder systems with limited access to chemical fertilizers (Sanchez, 2002), and was accurately assessed using the combined method of raw reflectance VNIRS + pH, EC, PR15 and PR45. The identification of C_{act} using the above combined method and OM_{LOI} by raw reflectance VNIRS therefore facilitates low cost and rapid

Table 1.6 Statistical results of cross-validation of each SQ indicator predicted by the best prediction model

	F	R2	RMSEP	RPD	Predictor Variables	
					VNIRS	Measured variables
Sand	18	0.85	5.91	2.56	Raw	pH,EC,PR15,PR45
Silt	15	0.74	4.95	1.95	Raw + first	none
Clay	10	0.87	3.35	2.80	Raw	none
OM_{LOI}	10	0.91	11.08	3.40	Raw + first	none
C_{act}	14	0.92	78.6	3.61	Raw	pH,EC,PR15,PR45
WSA	18	0.81	8.64	2.29	Raw	pH,EC,PR15,PR45, nutrient contents
Θ_{fc}	18	0.83	0.03	2.44	Raw	pH,EC,PR15,PR45, nutrient contents
Θ_{pwp}	14	0.89	0.02	2.98	Raw	none
AWC	n/a	0.67	0.02	1.74	Raw	pH,EC,PR15,PR45, nutrient contents
PR15	8	0.15	210.53	1.08	Raw	none
PR45	17	0.62	410.27	1.63	Raw + first	none
pH	19	0.77	0.29	2.05	Raw + first	none
EC	10	0.44	0.03	1.34	Raw	none
CEC	10	0.88	1.81	2.89	Raw + first	none
P	18	0.60	19.19	1.58	Raw + first	none
K	10	0.58	119.32	1.54	Raw	pH,EC,PR15,PR45
Ca	15	0.91	328.57	3.27	Raw	none
Mg	8	0.82	56.14	2.38	Raw	pH,EC,PR15,PR45
Cu	10	0.84	0.6	2.50	Raw + first	none
Zn	8	0.41	9.01	1.31	Raw	pH,EC,PR15,PR45
S	14	0.70	1.59	1.82	Raw	pH,EC,PR15,PR45

n = 227

Table 1.7 Interpretation of kappa statistics (Viera and Garrett, 2005).

κ	Level of Agreement
< 0	Less than chance agreement
0.01-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-0.99	Almost perfect agreement

Table. 1.8 Kappa statistics for determination of textural category and soil quality category by raw reflectance VNIRS-PLSR or best prediction model

	Raw reflectance VNIRS				**Best method**				
	Po[a]	Pc[b]	κ	Interpretation	Po[a]	Pc[b]	κ	Interpretation	Method
Clay	0.89	0.51	0.78	Substantial agreement	0.89	0.51	0.78	Substantial agreement	Raw
OM_{LOI}	0.84	0.34	0.76	Substantial agreement	0.85	0.34	0.77	Substantial agreement	Raw + first
Cact	0.81	0.35	0.70	Substantial agreement	0.87	0.36	0.79	Substantial agreement	Raw + pH,EC,PR15,PR45
WSA	0.50	0.36	0.21	Fair agreement	0.50	0.37	0.21	Fair agreement	Raw + pH,EC,PR15,PR45
AWC	0.59	0.36	0.36	Fair agreement	0.65	0.37	0.45	Moderate agreement	Raw + pH,EC,PR15,PR45, nutrient contents
PR15	0.98	0.98	0.00	Less than chance agreement	n/a	n/a	n/a	n/a	Measured
PR45	0.30	0.28	0.03	Slight agreement	n/a	n/a	n/a	n/a	Measured
pH	0.75	0.39	0.60	Moderate agreement	n/a	n/a	n/a	n/a	Measured
EC	0.86	0.76	0.42	Moderate agreement	n/a	n/a	n/a	n/a	Measured
P	0.65	0.41	0.40	Fair agreement	n/a	n/a	n/a	n/a	Measured
K	0.88	0.87	0.08	Slight agreement	n/a	n/a	n/a	n/a	Measured
CSQI	0.54	0.26	0.38	Fair agreement	0.71	0.26	0.61	Substantial agreement	n/a

[a] Po = percentage observed agreement (%)

[b] Pc = agreement by chance (%)

landscape wide assessments that can also account for high spatial variability by increasing the number of analyzed samples.

1.6. CONCLUSION

VNIRS-PLSR can successfully predict many SQ indicators for smallholder agricultural soils in Western Kenya across different parent materials within one soil order. In particular, several primary properties and correlated secondary properties showed substantial predictability (R^2 > 0.80; RPD > 2) including Ca, OM_{LOI}, C_{act}, Θ_{pwp}, CEC, clay, sand, and Cu, which were validated by both cross-validation of the leave-ten-out method and independent sample model validation. Of the spectral preprocessing methods that were compared, raw reflectance had the highest efficiency in our study. However, it is important to note that the accuracy of PLSR with raw reflectance spectra is primarily associated with reflectance measurements in controlled environments with constant illumination. The predictability for several SQ indicators were further improved by combining VNIRS with several field and laboratory measurements. As a result, using the best-combined prediction models, C_{act}, OM_{LOI}, Θ_{pwp}, Ca, CEC, clay, sand, Θ_{fc}, Cu, Mg, and WSA were predicted with sufficient accuracy.

The results from categorical analysis suggest that VNIRS could be substituted for selected laboratory analyses for assessing soil physical and biological quality constraints apart from soil compaction. The combined method of raw reflectance VNIRS and field measured SQ indicator values further improved accuracy, especially for C_{act}. The combined use of VNIRS and conventional laboratory analysis provides a comprehensive assessment of SQ indicators that

relate to key physical, biological, and chemical processes in soils. We conclude that these rapid and low cost methods have potential for use in identifying SQ constraints and associated yield limiting factors, may inform appropriate site-specific soil management practices, and could be used for SQ monitoring efforts.

1.7. ACKNOWLEDGEMENTS

This work was in part supported through a USDA-NIFA Special Grant on Computational Agriculture. Samples were derived from sites established and maintained through the Coupled Natural and Human Systems Program at Cornell, and the efforts of James Kinyangi, Solomon Ngoze, Joseph Kimetu, John Recha, Johannes Lehmann and many field staff members and smallholder farmers in Kenya. It was funded by the NSF Bio-Complexity Initiative (BCS-0215890). We also acknowledge support from the Bradfield family, the Saltonstall family, Cornell Graduate School, Heiwa Nakajima Foundation and the Mario Einaudi Center for International Studies, and from collaborators at the World Agroforestry Centre offices in Nairobi and Kisumu, and at the Kenya Forestry Research Institute in Maseno, Kenya. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

CHAPTER 2. COMPARISON OF SPECTROSCOPY, LANDSCAPE MODELING AND GEOSTATISTICS FOR SOIL ORGANIC CARBON ASSESSMENT IN A COFFEE AGROFORESTRY SYSTEM

2.1.ABSTRACT

Field-scale quantification of soil organic carbon (SOC) stocks generally requires large sample numbers in order to account for high spatial variability. The objective of this study was to evaluate the use of various techniques, including visible and near-infrared reflectance spectroscopy (VNIRS; 350-to-2500 nm; 1 nm resolution), landscape attributes, remote sensing, and geostatistics for SOC assessment in a tropical coffee agroforestry system within a micro-watershed in Costa Rica. VNIRS was employed on 478 air-dried samples of Andisols and 72 reference samples were additionally analyzed by conventional dry combustion. A partial least squares regression model using raw reflectance spectra successfully predicted SOC contents with R^2 of 0.82, RMSEP of 11.70 g kg^{-1}, and the ratio of performance to deviation (RPD) of 2.36 with the cross-validation and mean R^2 of 0.75, RMSEP of 12.87 g kg^{-1}, and RPD of 2.10 with multiple jackknifing. Pearson correlations between measured and VNIRS-PLSR predicted SOC contents, terrain attributes derived from a digital elevation model (DEM), and remote sensing-derived leaf area index were also assessed. A multiple linear regression (MLR) model with elevation and plan curvature explained only 14.8 % of the variability with an RMSEP of 27.10 g kg^{-1}. The stratification of the watershed into soil-landscape units (SLU) was necessary. Pearson correlations were only significant in three of six SLUs and were not sufficient for the use of

MLR models to whole watershed SOC stock assessment. The spatial structure of SOC was assessed with auto- and cross-variograms using the combinations of measured SOC, VNIRS-PLSR predicted SOC, and landscape attributes where the nugget:sill ratio showed moderate spatial structure. However, ordinary kriging and cokriging showed poor predictability of the SOC contents with only $0.20 < R^2 < 0.28$ and 23.16 g kg^{-1} < RMSEP < 24.62 g kg^{-1} due to localized variability. In conclusion, VNIRS has the greatest potential for estimating SOC stocks in Andisols of a tropical coffee agroforestry system, and soil-landscape modeling and geostatistics were of limited benefits due to spatially complex variability structures.

2.2. Keywords

Soil organic carbon; VNIR spectroscopy; PLSR; soil reflectance; agroforestry; watershed; soil-landscape modeling; kriging

2.3. Introduction

Coffee is an important agricultural commodity (Pendergrast, 2009), of which more than a third is produced in Latin America (Varangis et al., 2003). In Central America, substantial areas for coffee production are grown together with leguminous trees such as *Erythrina spp, Inga spp,* or *Gliricidia sepium* (Beer et al., 1998), originally as a way to reduce unfavorable climatic conditions and to improve the size and the quality of coffee beans (Muschler, 2001). This complex mixture also results in 1) increased organic residues, 2) reduced soil erosion, compaction, and runoff, and 3) a buffered microclimate (Fournier, 1988; Beer et al., 1998; Vaast

et al., 2007) that overall lead to both higher accumulation and more complex distribution of soil organic carbon (SOC) compared to monoculture systems (Payán et al., 2009).

SOC is a fundamental property for optimizing soil physical, chemical and biological quality (Magdoff and van Es, 2009) and is an important component of the global carbon cycle. Beneficial effects of agroforestry systems on minimizing the loss of SOC should be monitored and valued accurately, and therefore, a low cost, rapid, and reliable method to estimate the spatial distribution of SOC is required at a farm scale. Furthermore, this can contribute to a better understanding of carbon cycles within ecosystems and establishing the incentives for including soil carbon within carbon credit frameworks.

High spatial variability and non-linear temporal dynamics (Soussana et al., 2004) of SOC can partly be the result of land-use history, pedogenesis, differences in soil physical properties and ecosystem functions (Herrick and Whitford, 1995) that relate to complex topography, vegetation and agricultural management. The assessment therefore requires the analysis of large sample numbers, which is generally costly.

Visible and near-infrared reflectance spectroscopy (VNIRS) is an alternative to conventional analytical methods for assessing SOC and other soil properties in the context of precision agriculture (Viscarra Rossel and McBratney, 1998a; Thomasson et al., 2001; Bilgili et al., 2011), soil C monitoring (Post et al., 2001; Brunet et al., 2007), quantitative soil-landscape modeling (McKenzie et al., 2000), and digital soil mapping (Lamsal, 2009). VNIRS detects reflectance

characteristics caused by spectral overtones due to stretching and bending vibrations in molecular bonds such as C-C, C-H, N-H and O-H (Dalal and Henry, 1986), the size and shape of soil particles, the voids between soil particles, and the arrangement of soil particles (Reeves et al., 2011: in review). VNIRS employs a monochromator or sometimes Fourier reflectance spectra that can be interpreted using multivariate statistical or data mining techniques that relate spectra to directly-measured soil characteristics (Viscarra Rossel et al., 2006; Reeves et al., 2011: in review). The resulting prediction models can be used to assess soil properties of unknown samples.

Different statistical data processing methods may provide variable predictive outcomes. Some researchers used the first derivatives of soil reflectance spectra (Chang et al., 2001; Reeves et al., 2002; Shepherd and Walsh, 2002; Russell, 2003; Morón and Cozzolino, 2004; Couteaux et al., 2005), while others used raw spectra (Reeves et al., 2006; Bilgili et al., 2010), or second derivatives (Salgó et al., 1998; Fystro, 2002). Although Chang et al. (2001) argued that no universally perfect calibration methods exist, partial least squares regression (PLSR) is emerging as a successful method that has similar or better predictability than other methods (Viscarra Rossel et al., 2006; Bilgili et al., 2010). Using the PLSR model, Vasques et al. (2009) have successfully estimated SOC contents with $R^2 > 0.86$ using the visible and near infrared spectral range (350-2500 nm). Similar or better results such as $R^2 > 0.88$ were found by McCarty and Reeves (2006), and $R^2 > 0.84$ by Reeves et al. (2002) for SOC as well as $R^2 > 0.93$ by Kinoshita et al. for SOM (2011, in review).

Traditionally spatial estimation of SOC has been conducted by point measurements that are averaged and considered as representative of the site without accounting for significant local soil carbon variability (Eswaran et al., 1995). This type of extrapolation is known as the "measure and multiply" approach (Schimel and Potter, 1995). As an alternative, non-spatial extrapolation with quantitative soil-landscape models or spatial interpolation techniques are adopted for soil carbon estimation, often referred to as the "paint by numbers" (Schimel and Potter, 1995) approach. Soil-landscape modeling utilizes various environmental factors such as topography, hydrology, or geology that are related to soil carbon contents, and these relationships are used to upscale the point estimations of soil carbon to larger areas (Thompson and Kolka, 2005). A range of studies have demonstrated soil-landscape modeling in different locations (Moore et al., 1993; Chaplot et al., 2001; Florinsky et al., 2002; Terra et al., 2004; Thompson and Kolka, 2005) which resulted in explaining 20 to 88% of soil carbon variability using one to five terrain attributes. Thompson and Kolka (2005) stratified the watershed into different slope aspects based on the assumption of the presence of variable soil temperature and soil moisture contents, which gained model validation r^2 values of 0.020 to 0.802.

The use of VNIRS-derived data as an input for modeling spatial distribution of SOM also has not been widely studied to date (Lamsal, 2009) and none have been done in agroforestry systems. Spatial interpolation methods such as cokriging (COK; Ersahin, 2003) and regression kriging (RK; Hengl et al., 2004; Bilgili et al., 2011) can additionally be used with VNIRS data as covariates to estimate SOC for unsampled points. Kravchenko and Robertsen (2007) showed modest improvements in estimating soil C using a combined method of regression kriging and

topographical or yield information in a relatively flat arable field. Takata et al., (2007) showed a prediction error of 5.57 (Mg C ha^{-1}) in kriging soil carbon concentration using a combination of vegetation data acquired from MODIS satellite images as well as topographic data in agricultural fields. Bilgili et al. (2011) further explored the possibility of combining COK or RK with VNIRS to predict various soil properties on a 32 ha field, and gained higher accuracy in estimating SOM compared to ordinary kriging (OK).

Leaf Area Index (LAI) denotes the leaf area per soil area (Gower et al., 1999) and is a key indicator of plant physiological activity related to the carbon cycle through processes such as photosynthesis and net primary production. Normalized Difference Vegetation Index (NDVI) reflects LAI, thus the local productivity and, potentially, the litter inputs to the soil. In addition, LAI can indicate the canopy contribution to rainfall interception and litter production, which both contribute to soil carbon levels. Agroforestry creates heterogeneity inside the plots, for LAI but also for root distribution, both factors are assumed to influence the SOC distribution.

The combined use of geostatistics, VNIRS, landscape attributes, and remotely sensed NDVI is an attractive method for determining the spatial distribution of SOC in agroforestry systems because VNIRS allows for inexpensive analysis of many samples from a large area while geostatistics utilizes information on local variability and spatial dependence.

The objectives of this study were to 1) develop a VNIRS-PLSR model to estimate soil organic carbon (SOC) contents of a tropical Andisols within a 0.9 km^2 agroforestry watershed in Costa Rica, 2) analyze correlations between SOC, LAI, root distribution and landscape attributes for multiple linear regression (MLR) analysis, 3) analyze the spatial structure of SOC in the

watershed, and 4) assess the feasibility of combining VNIRS derived data and correlated secondary variables in geostatistics to estimate SOC.

2.4. Materials and methods

2.4.1. Location, climate and soil

The research site is located in the Central-Caribbean area of Costa Rica. It is a part of the Reventazón river basin on the slope of the Turrialba volcano, which eventually drains into the Caribbean Sea (Fig. 2.1). The research site is part of the Aquiares Coffee Farm (6.6 km^2), which is "Rainforest Alliance™" certified and it is a watershed of Mejias creek situated between −83◦ 44′ 39″ and −83◦ 43′ 35″ W, and between 9◦ 56′ 8″ and 9◦ 56′ 35″ N. The watershed has an area of 0.9 km^2 with an elevation ranging from 1020 to 1280 m a.s.l. and a mean slope of 11.31°. The climate is tropical humid with no dry season (Peel et al., 2007) and experiences strong influence from the Caribbean Sea. Between 1973-2009, the mean annual precipitation was 3014 mm in Aquiares. Gómez-Delgado et al. (2011) reported mean monthly net radiation between 5.7 and 13.0 MJ m^{-2} d^{-1}, air temperature between 17.0 and 20.8 °C, relative humidity between 83 and 91 % and windspeed at 2 m high from 0.4 to 1.6 m s^{-1} at the research site in 2009. Mora-Chinchilla (2000) described the geological characteristics of the basin as volcanic avalanche deposits that originated in the collapse of the southeastern slope of the Turrialba volcano. The soils are classified as Andisols (USDA Soil Taxonomy) and are characterized by high organic matter contents, very high infiltration capacities, and biological activities. The superficial runoff on plots is very low even on slopes, and the sediment production by the whole

watershed is very low, around 1 t ha^{-1} yr^{-1} (Gómez-Delgado, 2010). There is some presence of lava flows, agglomerates, lahars and ashes that are typical for volcanic regions.

The vegetation is a mixture of homogeneously planted coffee trees (*Coffea arabica L., var Caturra*) on bare soils that are shaded by widely spaced *Erythrina poeppigiana* trees. The coffee trees have grown to a 2.5 m canopy height, with average LAI≈3 and 20 % canopy openness from the initial density of 6300 plants ha^{-1} after 30 years from the initial planting. Every March, 20 % of the coffee trees are intensively and selectively pruned, and cut materials are removed from the site. The *Erythrina* trees have 20 m high canopies and that cover 12.3 % of the surface area at a density of 12.8 trees ha^{-1}.

2.4.2. Soil sampling

Soil samples were collected from the top 0-to-5 cm of the soil profile using a regular push-tube auger after the removal of the litter layer. The sampling depth was selected from our interest in soil C accumulation with variable overground vegetation densities, as Payán et al. (2009) found that the effects of shade trees on soil C was limited to top 0-to-5 cm when comparing the distance-to-trunk of 0 and 2 m.

Grid sampling was used after dividing the watershed into 23 sub-compartments based on MODIS satellite images of 230 m per pixel. Twelve transects were made in the E-W and six in the N-S direction within the MODIS grid with a sampling interval of 80 m and 40 m, respectively (Fig. 2.2). Higher sampling densities for the N-S transects were used to account for

Fig. 2.1 a) Location of Reventazón river basin in Costa Rica, Central America. b) Position of experimental basin within Reventazón river basin. c) The "Coffee-Flux" experimental basin in Aquiares farm and its experimental setup. (Gómez-Delgado et al., 2011)

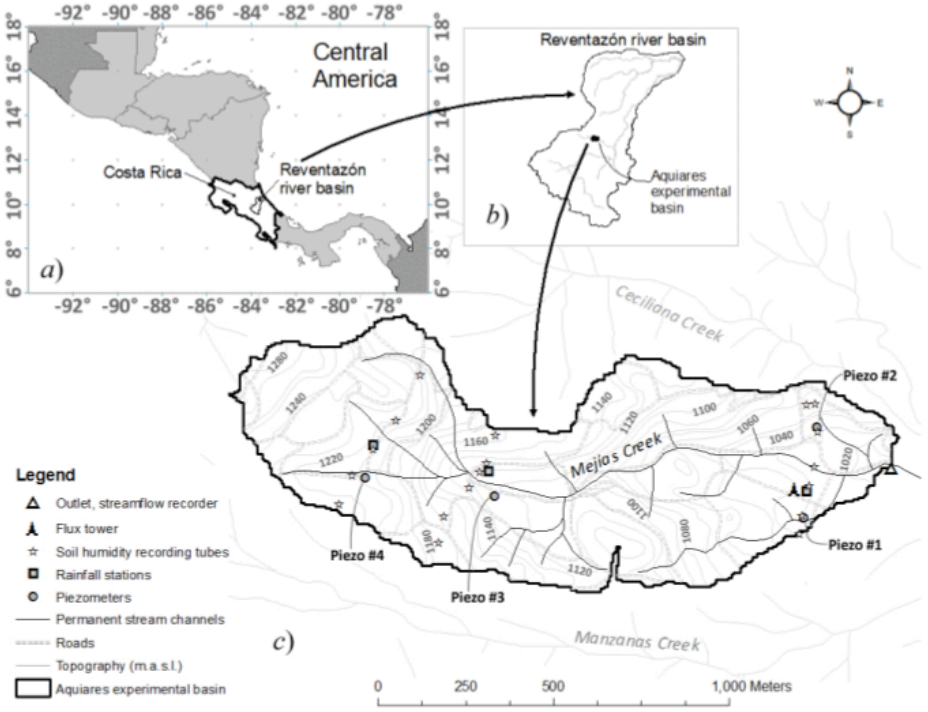
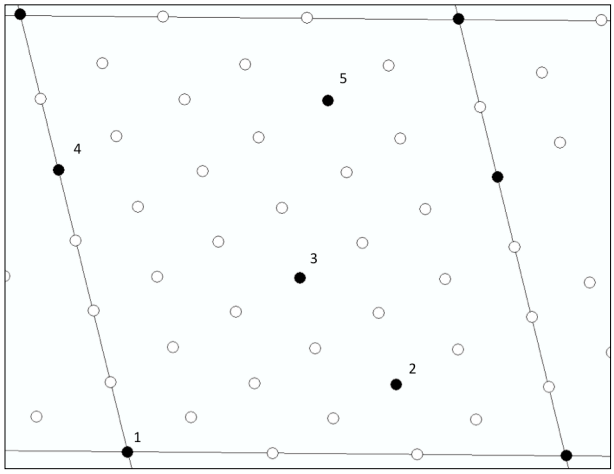


Fig. 2.2 MODIS grid and sampling points. The numbered points representing the reference set.



higher variability in topography. Five samples per MODIS grid (76 total) were denoted as the "reference set" that was later used for both conventional SOC analysis by dry combustion and VNIRS analysis (Fig. 2.2). The remaining 449 sample "prediction set" was collected only for VNIRS analysis after omitting some sampling points due to inaccessibility by watercourses, roads and dense vegetation along the rivers. At each sampling point, eight and four cores were taken for the reference set and prediction set for SOC, separated by a distance of about 2 m, which were thoroughly mixed in a plastic bag to obtain a representative sample. The collected soil samples were air-dried at 40°C for 72 hours and ground using mortar and pestle to pass through a 2 mm sieve. Visible roots were removed to avoid effects on the reflectance spectra.

2.4.3. Geomorphological characterization

Geomorphological properties of the study site were derived from a digital elevation model (DEM) created using the 5 m horizontal resolution and 10 m vertical resolution digital terrain model obtained from the TERRA-1998 project (scale ≈ 1: 25000; CENIGA, 1988). The geomorphological properties of elevation, slope, aspect, profile curvature, plan curvature, and specific catchment area (SCA) were derived using GRASS 6.4.1RC2 (GRASS Development Team, 2011). The compound topographic index (CTI) was also computed, which is the natural logarithm of the SCA to the local slope angle (tan β; in degrees) in each cell (Beven, 1997):

$$CTI = \ln (SCA / \tan \beta) \quad [1]$$

The CTI is also referred to as the steady state wetness index (Moore et al., 1991). SOC contents are influenced by higher soil water contents, lower heterotrophic decomposition, and the reduction of drought stress (Tate and Ross, 1997).

The topographic position index (TPI) was determined using the ArcGIS 9.3 (ESRI, Redlands, California) Topography Tools (Dilts, 2010), and was calculated as the difference between the cell elevation and the average elevation of the neighborhood cells (9 × 9 cells; Jenness, 2006). A positive TPI indicates that a cell has a higher elevation compared to surrounding cells. The combination of TPI and local slope can be used to categorize each cell into soil landscape units (SLU), which are valley, toeslope, footslope, backslope, shoulder, and upperslope. The SLU represents the distribution of soils over the landscape. Jian-Bing et al. (2006) showed statistically significant differences in SOC concentrations in different slope positions, where toeslope had the highest concentration followed by footslope, backslope and shoulder.

2.4.4. Leaf Area Index analysis

LAI was collected and presented by Taugourdeau et al. (2010), where detailed methods are described. Briefly, a satellite image acquired in March 2010 from Worldview2 (DigitalGlobe, 2009), with 2m resolution multispectral and 0.5 m resolution panchromatic images was utilized. Normalized Difference Vegetation Index (NDVI) was obtained as an index to measure vegetation cover. It is based on the fact that healthier vegetation reflects more near infrared radiation (ρNIR) and less red radiation (ρRED):

$$NDVI = (\rho NIR - \rho RED)/(\rho NIR + \rho RED) \quad [2]$$

Although the correlation between NDVI and LAI exists (Rouse et al., 1973), a specific calibration is necessary depending on the type of vegetation, the obtained image and the soil types (le Maire et al., 2006). Therefore, the effective LAI (LAI_{eff}) was measured using an LAI 2000 (LI-COR Environmental, Lincoln, NE) to help the calibration process, which was further corrected with actual LAI measured by collecting leaves of coffee trees (LAI_{TRUE}). The final LAI of the high-resolution satellite image (LAI_{HR}) was obtained as:

$$LAI_{HR} = \frac{a}{\ln (NDVI_{HR})} + b \qquad [3]$$

where a and b are parameters developed from the field measured LAI data.

2.4.5. Laboratory analyses

SOC as well as soil nitrogen were analyzed at USDA-ARS laboratory (Beltsville, MD) using a LECO TruSpec CN analyzer (LECO, St Joseph, Michigan). The dried and ground samples (0.6 g) were roller milled 24 hours prior to analysis, and two replicates were analyzed to observe instrumental errors and were generally less than ± 0.5 %. Soil texture was assessed at the Cornell Soil Health Testing Laboratory (Ithaca, NY) using a rapid quantitative method developed by Kettler et al., (2001). The soil sample was dispersed with 3% hexametaphosphate ($(NaPO_3)n$) and a combination of sieving and sedimentation steps was used to separate size fractions. The

sand content of the soil samples with SOC > 5 % were overestimated due to the presence of agglomerate and they were excluded for further analyses.

2.4.6. Visible near infrared spectroscopy (VNIRS)

The reflectance of soil samples was determined in both the visible and near infrared spectral regions between 350 nm and 2500 nm at 1 nm resolution using a Fieldspec Pro hyperspectral sensor (Analytical Spectral Devices, INC., Boulder, CO). Air-dried samples were placed in a 4 cm diameter optical quality petri dish, and spectral reflectance was collected through the glass bottom at a constant angle (55 degrees from horizontal) from a distance of 4 cm, in an enclosed box. A tungsten quartz halogen lamp was used as a light source (Muglight sensor attachment; Analytical Spectral Devices, INC., Boulder, CO), and the reflectance was measured using a fiber optic cable. After 50 readings, the sample petri dish was rotated 90 degrees and another 50 readings were taken, which were averaged to produce two spectra per sample. The signal was optimized before the measurement and the accuracy of the detector was checked every 8 samples using a white Spectralon standard. When the white Spectralon reading was not stable, the instrument was recalibrated.

2.4.7. Spectral preprocessing

Reflectance data were exported from binary to an ASCII file using ViewSpec Pro software (Analytical Spectral Devices, INC., Boulder, CO). The two spectral readings per sample obtained during measurement were averaged using Unscrambler 10.0.1 (CAMO software, Oslo,

Norway, 2010). Derivative values were used to overcome baseline noise and to amplify spectral signals (Reeves et al., 2002). First derivatives were calculated using the Savitsky-Golay transformation (Savitzky and Golay, 1964), which performs a second order polynomial regression using 4 points to the left and one point to the right side of each point for smoothing. This transformation method is suitable for VNIRS applications since it removes noise caused by ambient illumination differences from the data while retaining original spectral characteristics with minimum distortion (Ruffin and King, 1999).

2.4.8. VNIRS modeling

The VNIRS analysis utilizes empirical models based on correlations between soil reflectance and soil property values. Calibration models for predicting SOC values were constructed utilizing the partial least squares regression (PLSR) method (Bilgili et al., 2010; Kusumo et al., 2010; Hively et al., 2011). It relates two variables, X (spectral readings) and Y (measured soil property values), by a linear multivariate model. Orthogonal and weighted linear combinations of the spectral readings are used for predicting each Y variable. PLSR is suited to handle data with strong collinearity in the (X) variables, which are usually more numerous than the observations (Y) that they predict. The selection of the number of factors (F) to include was critical to avoid over-fitting or under-fitting of the model to the data. Over-fitting of calibration models reduces their ability to predict soil properties of new unknown soil samples. In this study the smallest number of F that produces sufficiently small root mean square error of prediction (RMSEP) was selected by analyzing the total residual variance plot. Also, the maximum F was restricted to 20. The validity of the VNIRS-PLSR method was determined by two general procedures, cross

validation and independent calibration-validation. For cross validation using the leave-one-out method (Kooistra et al., 2001; Reeves et al., 2002; McCarty and Reeves, 2006), both the raw and the first derivative data of the reference sample spectra (n=76) were used. Spectral outliers were identified with Hotelling T^2 values, equivalent to the squared Mahalanobis distance H (McLachlan, 1992), which determines the distance of a spectrum from the average spectrum (Peltre et al., 2011). Any samples with Hotelling T^2 exceeding the critical limit at the p value of 0.05 were removed from the set and the cross validation was repeated.

Since our reference sample set was relatively small after the removal of four spectral outliers (n = 72), further analysis was required to validate the feasibility of the VNIRS-PLSR method. The reference sample set was randomly divided into two subsets containing 70 and 30 percent of the samples (n = 50 and n = 22) and the former sample set (n =50) was used as a calibration subset to predict the soil properties of the latter subset (n = 22). The random selection of the calibration and validation set was repeated six times to overcome the influence of outliers and clustering effects in the sample set. This method is known as a multiple jackknifing procedure, which allows an independent assessment of the prediction quality or uncertainty in a small data set (Bishop and McBratney, 2001).

Following the validation of the VNIRS-PLSR method, soil properties of the prediction set (n = 449) were analyzed using the cross validated model. Spectral outliers in the prediction set were identified in a similar manner to the cross validation method utilizing Hotelling T^2 values (p-value < 0.05).

2.4.9. Multiple Linear Regression

Simple exploratory data analysis was undertaken to distinguish important landscape attributes (geomorphological and LAI) governing SOC distribution in the micro-watershed of Aquiares. Pearson correlation coefficients were calculated between the variables for both the reference set (n = 72) and the combination of the reference set and the prediction set (n = 478) to elucidate significant correlations. Selected landscape attributes in each of the sample set were utilized to build multiple linear regression models and validated with the cross-validation with the leave-one-out method. Model assumptions of lack of multicollinearity, equal error variance, and normal and random residuals were assessed.

2.4.10. Geostatistical analyses

Ordinary kriging (OK) was employed to analyze SOC contents at unvisited locations within the study site. First, the reference set (n = 72) after the removal of spectral outliers was used to build a variogram that represents spatial correlation and covariance structure between data points (Burgess and Webster, 1980).

$$\hat{\gamma}(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} [z(x_i + h) - z(x_i)]^2 \qquad [4]$$

where $\hat{\gamma}(h)$ is the semivariance between two sampling points, $z(x_i)$ and $z(x_i + h)$, that are separated by a distance h, and m(h) is the number of pairs at h. Anisotropy was examined for the watershed due to the complex topography. Exponential and spherical models were fitted to both

isotropic and anisotropic variograms. Anisotropic surfaces were used to find the principal axis for determining the anisotropic variogram model (Isaaks and Srivastava, 1989). The nugget:sill ratio was utilized as an indicator of spatial dependence of SOC as well as selecting the most suitable model for the study site. OK was then applied as the geostatistical method for interpolation of SOC. The prediction is based on the weighting of the distances from neighboring samples to the target, which is governed by the computed variogram and calculated by:

$$Z^{*}_{OK}(X_0) = \sum_{i=1}^{n} w_i \, z(x_i) \qquad [5]$$

where $Z^{*}_{OK}(x_0)$ is the OK estimate at an unsampled location (x_0), n is the number of samples in the search neighborhood, and w_i is the weight assigned to the ith observation $z(x_i)$. Weights are assigned to each sample to minimize kriging variance, $E[Z^{*}(x_0) - Z(x_0)]$, and also to gain unbiased estimation (Webster and Oliver, 2007).

The prediction accuracy of the models was assessed by the coefficient of determination (R^2) and root mean square error of prediction (RMSEP) using cross validation of the leave-one-out method.

2.4.11. Cokriging

Cokriging (COK) incorporates one or more secondary variables, or covariables (Z_2), which are spatially correlated with laboratory measured SOC. The equation for COK is given by (Journel and Huijbregts, 1978; Matheron, 1978):

$$Z^*_{COK}(\mathbf{x_0}) = \sum_{i=1}^{n1} w_i z_1(\mathbf{x}_i) + \sum_{j=1}^{n2} w_j z_2(\mathbf{x}_j) \qquad [6]$$

where $Z^*_{COK}(\mathbf{x_0})$ is the COK estimate at an unsampled location (x_0), w_i and w_j are COK weights associated with the primary variable $z_1(x_i)$ and the secondary variable $z_2(x_j)$ at ith and jth locations, respectively, which are based on the cross-variogram:

$$\hat{\gamma}(\mathrm{h}) = \frac{1}{2m(\mathrm{h})} \sum_{i=1}^{m(\mathrm{h})} [z_1(\mathbf{x}_i + \mathrm{h}) - z_1(\mathbf{x}_i)] [z_2(\mathbf{x}_j + \mathrm{h}) - z_2(\mathbf{x}_j)] \qquad [7]$$

Cross-variograms were fitted using the linear model of coregionalization (LMC), which ensures that the two auto-variograms for the primary variables and one cross-variogram for the covariable have the same range and spatial structure (e.g. spherical or exponential function). The LMC ensures that the co-kriging system is positive-definite (i.e. all possible combinations of random variables have a positive variance), which is a prerequisite for COK (Goovaerts 1997).

In this study, the primary variables (Z;Z_1) were the laboratory measurements of the reference set (n = 72) and the covariables (Z_2) were the estimates from VNIRS-PLSR of the prediction set (n = 406).

2.4.12. Prediction accuracy

The accuracy of each PLSR model for SOC prediction, multiple linear regression, OK, and COK were evaluated using the coefficient of determination (R^2) of measured and model-predicted values. Root mean square error of prediction (RMSEP), a measure of precision calculated as the differences between model predicted values and observed values, was determined as in:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(Y_{pred.} - Y_{meas.})^2}{n-1}} \qquad [8]$$

The ratio of performance to deviation (RPD) was also calculated for determining the accuracy of VNIRS-PLSR models, which represents the ratio of standard deviation to standard error of prediction in each sample. It is a measure of the ability of a VNIRS-based model to predict SOC (Williams and Sobering, 1993).

2.5. Results and discussions

2.5.1. Soil properties

Soil texture was assessed by a rapid quantitative method and the research site was categorized in either loam or sandy loam textural classes. Summary statistics of SOC and soil N in 72 reference samples after the removal of four spectral outliers are shown on Table 2.1. The spatial variability

of SOC concentration was high, ranging from 45.0 g kg^{-1} to 154 g kg^{-1} across the watershed. Conant et al. (2003) measured coefficient of variation (CV) values of SOC within a microplot of 2 m × 5 m of 10 to 15 % in a cultivated agricultural site and 60 to 126 % in a forest site in Tennessee and Washington, compared to 32 % in our study. The shallow sampling depth of 0-to-5 cm may be related to high spatial variability in soil C contents (Conant et al., 2003) due to the influences from above vegetation (Payán et al., 2009), microclimate and soil physical properties (Saetre, 1999).

An average SOC of 87.0 g kg^{-1} is high, although Chesworth (2008) noted that Andisols commonly have soil C contents of more than 60 g kg^{-1} in both A and B horizons and in some humid conditions soil C contents can reach 200 g kg^{-1}. In addition, Nanzyo et al. (1993) have shown that even allophanic Andisols contain up to 150 g kg^{-1} SOC, whereas nonallophanic Andisols accumulate up to 230 g kg^{-1} SOC. The C:N ratio was relatively constant across the watershed (CV = 6 %) with a mean ratio of 10.9. This is in line with similar studies of coffee agroforestry and pasture systems on Andisols, which had a mean C:N ratio of 11.6 (Hoyos et al., 2005). The narrow range for C:N ratio is thought to originate from the homogeneous crop type and small contribution of leaf litter from leguminous shade trees.

2.5.2. Soil reflectance

Typical of soils, the reflectance was lower in the visible range and increased in the near-infrared range between 700 nm and 1800 nm (Fig. 2.3a, c). There were three distinctive absorbance bands at 1400, 1900 and 2200 nm. The spectral band at 1400 nm is strongly affected by OH features of free water and clay lattices whereas the band at 1900 nm is affected by OH features of free water

Table 2.1 Descriptive statistics of reference set and prediction set

	n	Min	Max	Mean	St.dev	CV[b]
SOC[a] (g kg^{-1})	72	45.0	154	87.0	27.6	32
SOC_{VNIRS_PLS} (g kg^{-1})	406	6.40	160	80.4	26.7	33
N (g kg^{-1})	72	4.60	12.9	7.90	2.20	28
C:N ratio	72	9.64	13.4	10.9	0.67	0.06

[a] SOC = soil organic carbon by dry combustion

[b] CV = coefficient of variation

and 2200 nm by OH features of clay lattices only (Hunt, 1980). The albedo of the reference set was 38% at the maximum (Fig. 2.3a) and was lower in comparison to other soil types such as in Ultisols (60 %; Kinoshita et al., 2011, in review). Higher SOM contents exhibit lower reflectance (Stoner and Baumgardner, 1981; Ben-Dor et al., 1999; Vagen et al., 2006), which may explain the relatively low reflectance of our dataset with SOC of up to 150 g kg^{-1}. Additionally, coarse soil texture is known to have low reflectance as seen in our samples of high sand content (Viscarra Rossel and McBratney, 1998a, b; Reeves et al., 2011, in review).

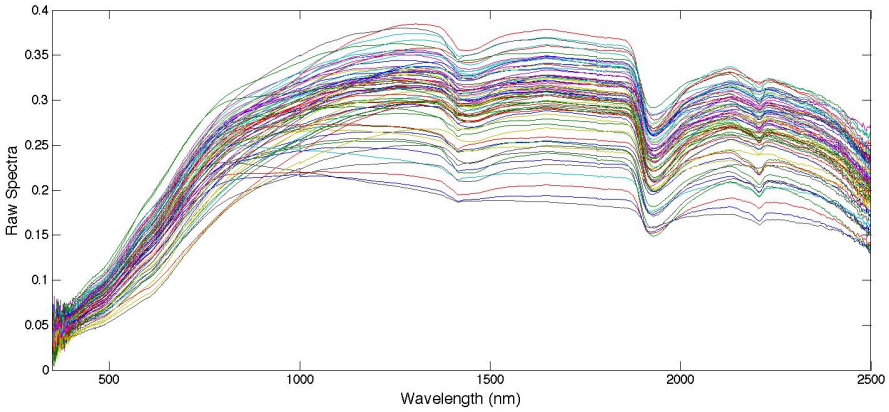
2.5.3. Prediction of soil properties by VNIRS

Due to the limited number of reference soil samples (n=72), the leave-one-out cross-validation method was adopted for the PLSR analysis. Cross-validation R^2 and RMSEP were compared for raw reflectance and first derivative reflectance (Fig. 2.3a, b) after the removal of four spectral outliers. The raw reflectance spectra showed higher R^2 values for SOC estimations compared to the first derivative spectra (0.82 vs. 0.58; Table 2.2) and lower RMSEP values (11.7 vs. 18.0 g kg^{-1}). This is in agreement with Bilgili et al. (2010), who found better predictability with raw reflectance for SOM and soil texture, though the reduction in the predictability with first derivative reflectance was more pronounced in our study. The need for utilizing first derivative reflectance was minimal in this experimental setup, because ambient illumination levels were kept constant during the spectral scanning, and the average of 100 spectral readings from two different positions were taken per sample.

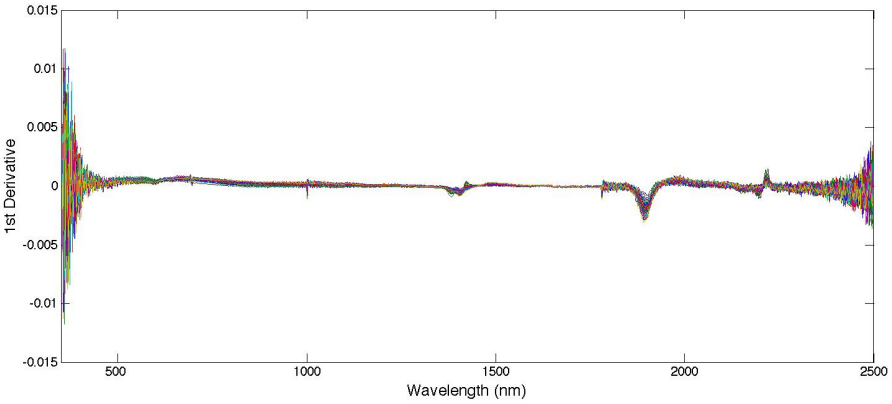
Comparison between raw reflectance VNIRS-PLSR predicted SOC values and measured values show points along the 1:1 line (Fig. 2.4).

Fig. 2.3 (a) Soil raw spectra of reference set (n = 72). (b) Soil first derivative spectra of reference set (n = 72). (c) Soil raw spectra of prediction set (n= 406)

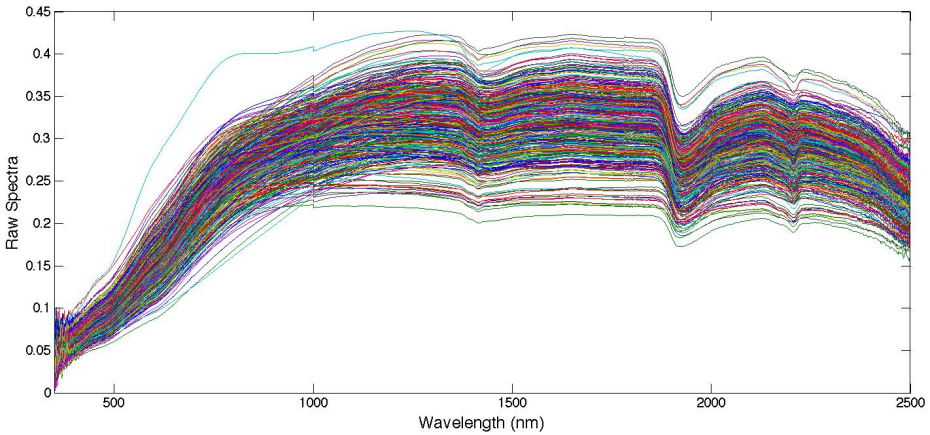
a)



b)



c)


Raw Spectra
Wavelength (nm)
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
500
1000
1500
2000
2500

Model accuracy was also compared using RPD since RMSEP alone cannot provide sufficient information on model predictability due to variable SD of the soil properties. In soil science, RPD > 2 is denoted as satisfactory (Chang et al., 2001). The RPD of 2.36 was obtained using raw reflectance compared to 1.68 by first derivative reflectance in our study, which shows the adequacy of the raw reflectance model to be further applied in estimating SOC of new unknown soil samples. Furthermore, multiple jackknifing validation was applied to be more conservative in the model validation since the cross-validation method tends to overestimate goodness of fit. The results show slightly lower predictability with mean R^2 of 0.75 and RMSEP of 12.9 g kg^{-1}. The RPD value ranged from 1.55 to 2.62 with the mean value of 2.10, which satisfies the threshold of 2 overall (Table 2.3). The poor calibration and validation in several data subsets can be attributed to spectral outliers within the Hotelling T^2 limit or clustering of spectral characteristics either in calibration or validation due to small sample numbers. The cross-validated raw reflectance model was applied to predict SOC contents of the prediction set (n = 406) after the removal of forty-three spectral outliers. The predicted SOC contents had mean values of 80.4 g kg^{-1} and a CV of 33 % (Table 2.1).

2.5.4. Multiple linear regression with landscape attributes

The correlation between SOC and various landscape attributes in the reference set (n = 72) were assessed using Pearson correlation coefficients. SOC had significant correlations ($\alpha = 0.05$) to both elevation (r = 0.380) and plan curvature (r = -0.337; Table 2.4). Higher elevation has shown higher SOC contents in various study sites. Thompson and Kolka (2005) showed that higher SOC contents correlate with higher elevation in a watershed in

Table 2.2 Statistical results of cross-validation using the leave-one-out of VNIRS[a] with PLSR[b] using raw and first derivative reflectance spectra for SOC (n = 72)

	F[c]	R^2	RMSEP[d]	RPD[e]
Raw spectra	6	0.82	11.7	2.36
First derivative spectra	6	0.58	18.0	1.68

[a] VNIRS = visible and near-infreared spectroscopy

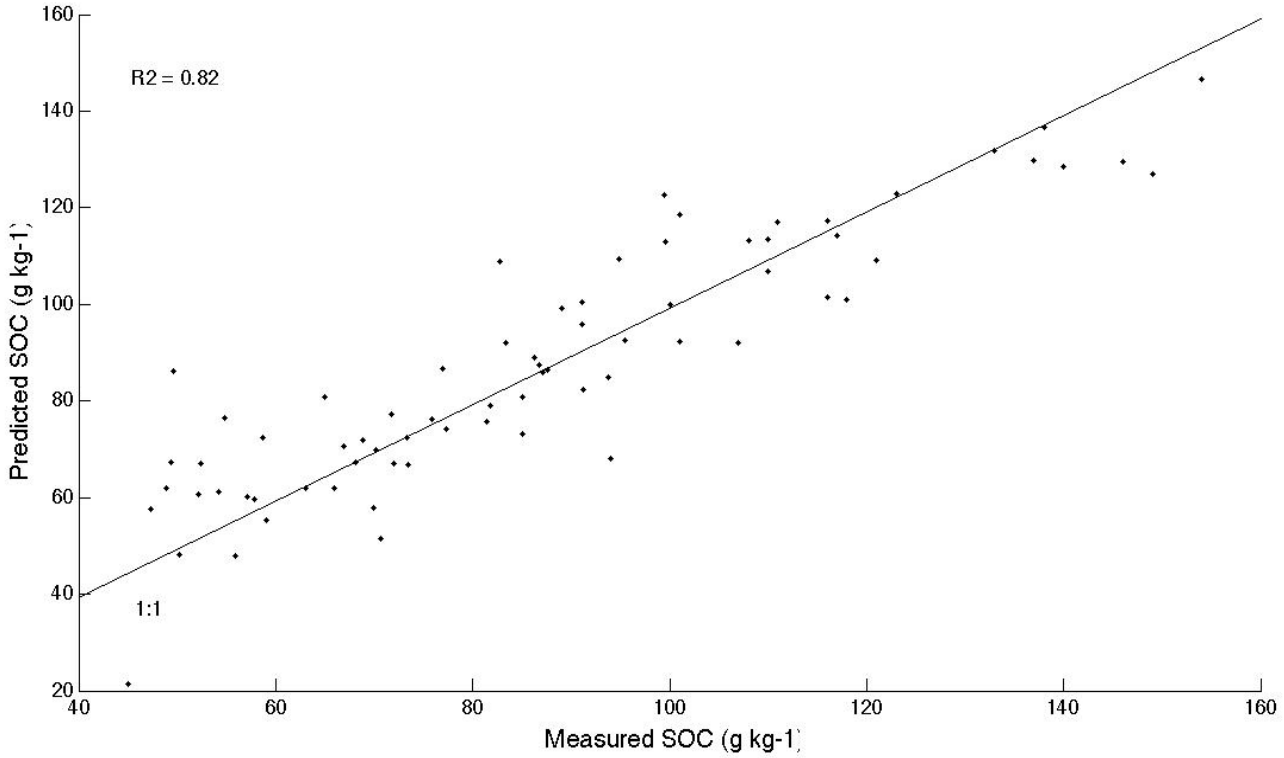
[b] PLSR = patial least squares regression

[c] F = number of factors used in PLSR

[d] RMSEP = root mean square error of prediction

[e] RPD = ratio of performance de deviation

Fig. 2.4 Comparison of measured SOC values by combustion and predicted SOC values by VNIRS-PLSR for reference samples (n = 72)



Kentucky, USA. Garten et al. (1999) and Bolstad and Vose (2001) also showed higher SOC contents over a range of 1000 m and concluded that cooler soil temperatures at higher elevation are the primary cause. Significant negative correlation between profile curvature and SOC shows higher SOC in concave topography, similar to Gessler et al. (2000). Convex topography is generally associated with convergent water flow and higher sedimentation, which accumulates more SOC. The multiple linear regression model explains 14.8 % of the variability in SOC using the two correlated landscape attributes and the mean RMSEP of cross-validation of the leave-one-out method was 27.1 g kg^{-1}.

The combined method of VNIRS-PLSR and laboratory-measured SOC values for landscape modeling was also undertaken. The mean SOC content of the study site was 81.4 g kg^{-1} when the reference set (n =72) and prediction set (n = 406) are combined (n = 478; Table 2.5). Pearson correlation coefficients between SOC and landscape attributes for the combination of reference and prediction set were low and statistically non-significant (Table 2.7). This decrease in correlations compared to the reference set (Table 2.4) can be attributed to significant local variability in SOC that masks the global trend within the micro-watershed. Consequently, soil-landscape unit (SLU) was used to stratify the study area to observe variable distribution of SOC in the study site (Fig. 2.5). The highest accumulation of SOC was found in toeslope areas (88.2 g kg^{-1}) and the lowest in valley positions (67.5 g kg^{-1}; Table 2.6), which were statistically different ($\alpha = 0.05$) using the Tukey test. This is in line with the study by Terra et al. (2004), who found higher SOC contents with lower slope gradients since erosion is minimal. In this study site, toeslope had the lowest mean slope gradient of 3.23° compared to the global mean of 13.00° (Table 2.6).

Table 2.3 Results of modified jackknifing for VNIRS-PLSR assessment of SOC

	Calibration set (n = 50)				Validation set (n = 22)		
Sequence	F	*R2*	RMSEP	RPD	*R2*	RMSEP	RPD
1	6	0.77	13.9	2.09	0.85	8.47	2.62
2	4	0.80	13.2	2.21	0.56	14.9	1.55
3	6	0.84	10.9	2.52	0.78	13.1	2.17
4	5	0.75	13.5	1.97	0.77	14.0	2.13
5	6	0.83	12.1	2.43	0.73	11.6	1.96
6	4	0.71	13.5	1.84	0.78	15.2	2.16
Mean	5.2	0.78	12.9	2.18	0.75	12.9	2.10

Table 2.4 Pearson correlation between measured SOC and terrain attributes from a DEM[a]

Terrain attribute	Correlation coefficient
Elevation	0.380*
Slope	-0.179
Aspect	0.084
Profile curvature	-0.107
Plan curvature	-0.337*
SCA[b]	-0.163
CTI[c]	-0.048
LAI[d]	0.065

* Significant at $\alpha = 0.05$

[a] DEM = digital elevation model

[b] SCA = specific catchment area

[c] CTI = compound topographic index

[d] LAI = leaf area index

Geomorphological properties in the watershed presented high variability (Fig. 2.6). Elevation ranged about 280 m and the slope from 0° to 40°, which represent the complex topography of the study site. The density of surface vegetation also varied significantly with LAI's of 0 to 8.2.

The stratification of the study site by SLUs improved correlation coefficients between SOC and the landscape attributes for some. In footslope positions, plan curvature had a significant (α = 0.05) correlation (r = -0.272), which indicates higher SOC in concave topography at the foot of a slope, similar to Gessler et al. (2000). Convex topography generally creates convergent flow and higher sedimentation, which cause SOC accumulations.

In upper slope positions, elevation was significantly correlated (α = 0.05) to SOC (r = -0.265) and this was contrary to the finding with the reference set. However, the high slope gradient (16.07°; Table 2.6) in this SLU can cause downward transportation of SOM in the topsoil due to erosion, and create sedimentation and lower C decomposition due to more likely anaerobic conditions in lower elevations (Jian-Bing et al., 2006). LAI was positively correlated to SOC in upper slope position (r =0.29), and this may be related to a significant correlation (α = 0.05) between elevation and LAI (r = -0.28) in the unit. In the valley position, elevation was again significantly correlated to SOC (r = -0.53) where the mean slope gradient was the highest in all the SLUs (Table 2.6).

In the study site, significant correlations between SOC and various landscape attributes were only found in footslope, upperslope, and valley positions, which account for only 29 % of the total area. The large local variability in SOC at each sampling point can be caused by many factors including variable pruning management and inherent heterogeneity in soil properties and volcanic parent material.

Table 2.5 Descriptive statistics of combined (measured and VNIRS predicted) SOC values in different soil landscape units

	Sample set (n=478)					
	n	Min	Max	Mean	St.dev	CV
All	478	6.42	160	81.4	26.9	33
Toeslope	66	22.3	160	88.2	28.3	32
Backslope	240	12.5	154	79.6	25.3	32
Shoulder	25	31.9	135	83.3	31.5	38
Footslope	66	18.0	140	82.1	24.9	30
Upperslope	60	6.42	137	84.3	30.0	36
Valley	21	21.1	123	67.5	27.8	41

Table 2.6 Descriptive statistics of various terrain attributes and LAI in each soil landscape unit

	Sample set (n=478)					
	n	Elevation (m)	Slope (°)	SCA (m^2)	CTI	LAI
All	478	1124	13.00	227	6.23	2.44
Toeslope	66	1135	3.23	223	7.54	2.45
Backslope	240	1123	13.08	64	5.98	2.39
Shoulder	25	1149	18.40	33	4.53	2.15
Footslope	66	1105	15.87	836	7.21	2.71
Upperslope	60	1140	16.07	16	5.01	2.39
Valley	21	1101	18.76	1014	7.56	2.53

Table 2.7 Pearson correlation between combined (measured and VNIRS predicted) SOC and various terrain attributes from a DEM

	Soil Landscape Unit						
Terrain attribute	All	Toeslope	Backslope	Shoulder	Footslope	Upperslope	Valley
Elevation	-0.04	0.078	0.032	0.219	-0.13	-0.265*	-0.532*
Slope	-0.045	-0.167	0.004	0.004	0.021	-0.1	-0.131
Aspect	0.011	-0.1	0.025	-0.078	0.072	0.12	-0.091
Profile curvature	0.038	0.099	0.044	-0.006	0.033	-0.006	-0.269
Plan curvature	-0.057	0.034	-0.016	0.016	-0.272*	-0.085	-0.197
SCA	-0.054	-0.121	-0.047	-0.233	-0.237	0.023	-0.041
CTI	-0.055	0.066	-0.059	-0.107	-0.205	0.044	-0.123
LAI	0.087	0.161	0.059	-0.27	0.216	0.29*	-0.092

Fig. 2.5 Soil landscape units within the experimental basin

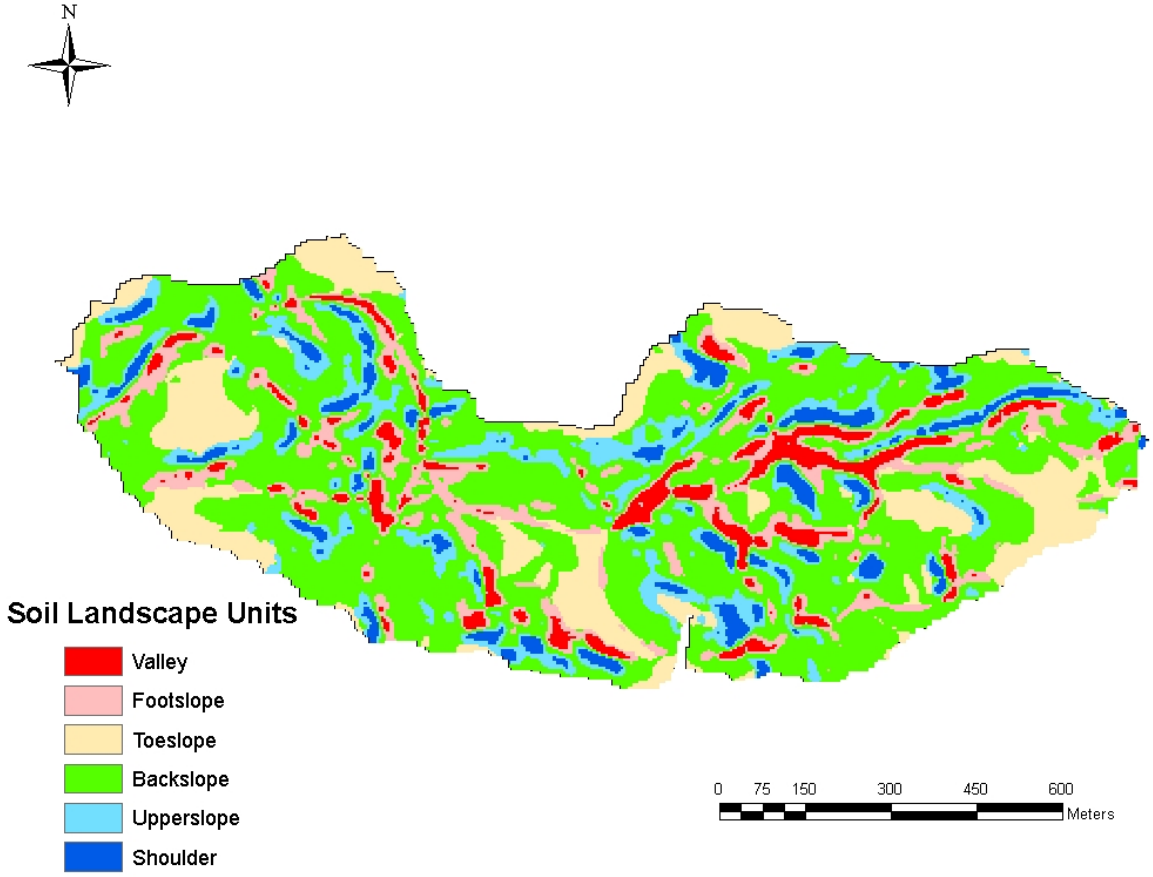
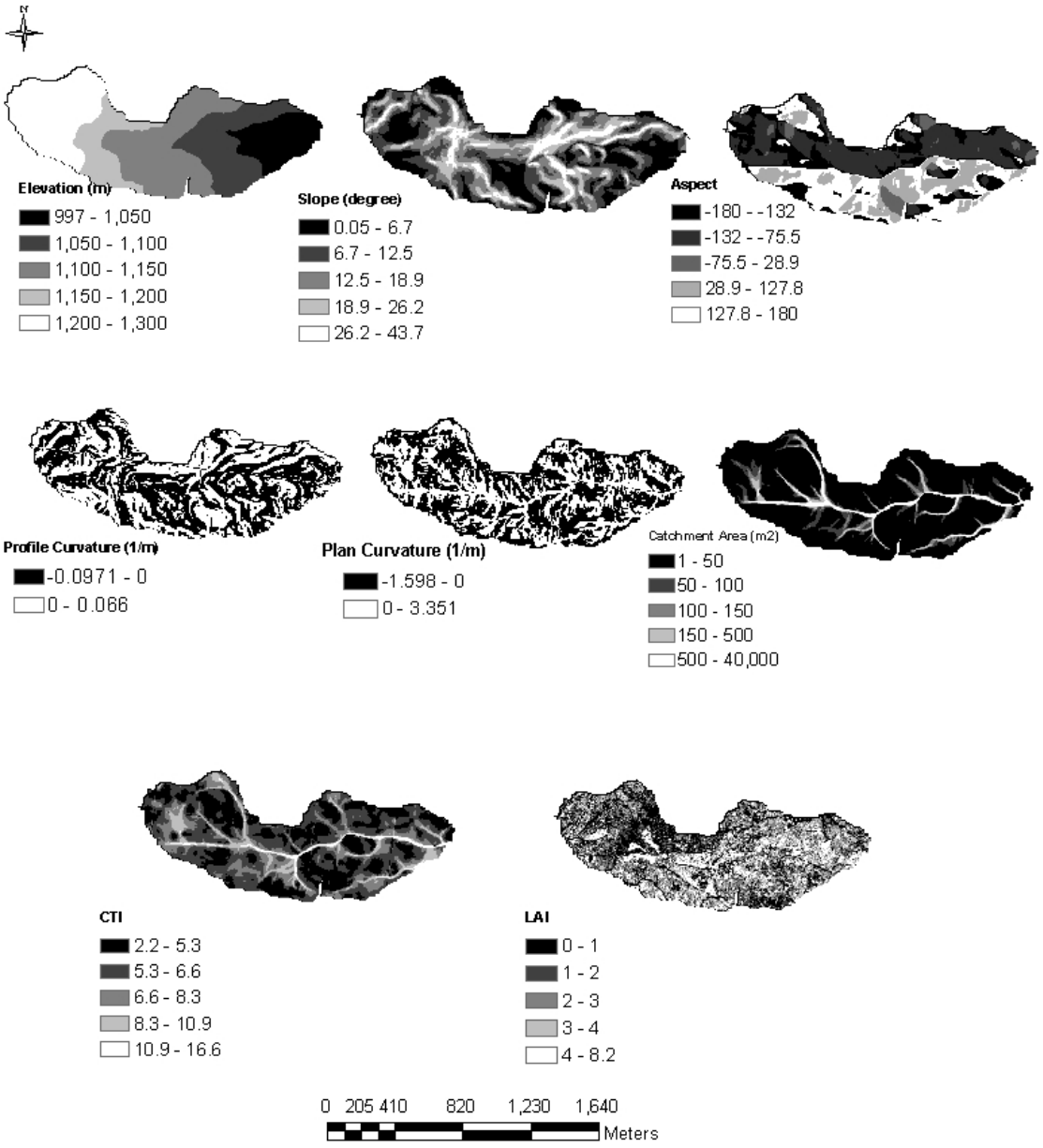


Fig. 2.6 Surfaces geomorphological properties in the experimental basin



2.5.5. Variogram models

Auto- and cross-variograms of the 72 reference samples for SOC were computed and modeled. Exponential and spherical models with or without anisotropy were fitted to the reference set of which the exponential model without anisotropy showed the best fit in sample variations (Table 2.8). The ratio of nugget to sill represents the spatial autocorrelation and the value of < 0.25, 0.25-0.75, and > 0.75 show strong, moderate and weak correlation, respectively (Cambardella et al., 1994). The nugget:sill ratio of our study ranged from 0.34 to 0.49 which indicates moderate spatial correlation of SOC. The interpolated surface has SOC ranging from 45.0 g kg^{-1} to 154 g kg^{-1} (Fig. 2.8).

2.5.6. Ordinary kriging vs. VNIRS

OK is a technique often used in precision agriculture to obtain soil properties at unsampled locations from measured values relying on spatial correlations (Webster and Oliver, 2007). This differs from VNIRS-PLSR prediction methods since it does not require any sample collections at the predicted locations.

The predictability of OK was compared to VNIRS using cross-validation of the leave-one-out method and the RMSEPs were 23.5 and 11.7, respectively (Tables 2.8 and 2.2). The coefficients of correlation were 0.26 and 0.82, respectively, indicating considerably better predictability with the VNIRS-PLSR approach. This can be partly related to the high nugget effect due to local variability of SOC in our study site with extremely complex topography as well as varying

surface vegetation densities. Variability of SOC contents in tree-based ecosystems is generally significantly higher than agricultural fields and is more difficult to represent by geostatistical methods (Conant et al., 2003). Nevertheless, it was relatively well predicted by the VNIRS-PLSR method, because it relies on soil reflectance characteristics and is not affected by unexplained spatial variability.

2.5.7. Cokriging

The high variability of SOC in this micro-watershed is thought to primarily originate from topography and vegetation (Terra et al., 2004; Payán et al., 2009). Therefore, higher sampling sizes combined with rapid analysis of SOC by VNIRS or the integration of landscape attributes as covariables may be superior to OK.

The correlation coefficients and RMSEP of the measured and estimated SOC contents of the reference set were again assessed using the leave-one-out cross-validation method and compared to the results obtained by VNIRS-PLSR and OK. Contrary to the findings in an agricultural field (Bilgili et al., 2011), the use of VNIRS lowered the predictability of SOC slightly compared to OK. The RMSEP was increased from 23.5 g kg^{-1} to 23.7 g kg^{-1} using the spherical model without anisotropy (Table 2.9).

Fig. 2.7 Interpolated surface of combined (measured and VNIRS predicted) SOC in Aquiares

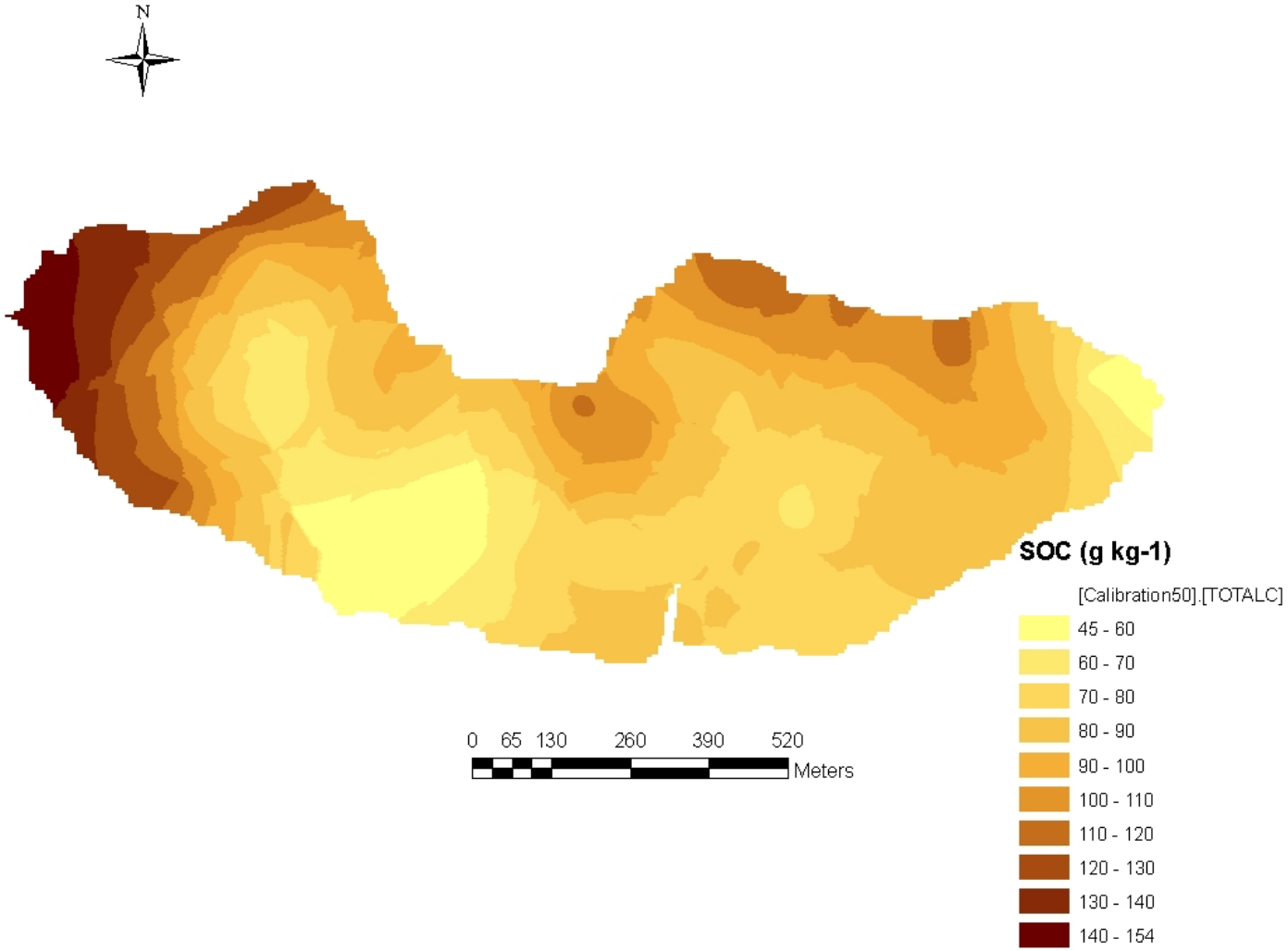


Table 2.8 Auto and cross-variogram model parameters of SOC data of reference set (n =72). c_0 nugget, c sill, and a range (m)

	OK[a]				COK[b]-VNIRS				COK-LAND[c]			
	c_0[d]	c[e]	a[f]	c_0/c	c_0	c	a	c_0/c	c_0	c	a	c_0/c
Exponential-anisotropy	28.4	84.3	992	0.34	28.5	84.2	992	0.34	32.7	88.3	992	0.37
Exponential	32.7	88.3	992	0.37	32.8	88.2	992	0.37	32.7	88.3	992	0.37
Spherical-anisotropy	37.6	85.5	992	0.44	37.5	85.4	992	0.44	41.9	88.2	992	0.48
Spherical	39.2	83.7	754	0.47	38.5	83.1	726	0.46	43.5	88.7	992	0.49

[a] OK = ordinary kriging

[b] COK = cokriging

[c] LAND = landscape attributes

[d] c_0 = nugget

[e] c = sill

[f] a = range (m)

Table 2.9 Comparison of OK and COK including VNIRS and landscape attributes.

	OK				COK-VNIRS				COK-LAND			
	EA[a]	E[b]	SA[c]	S[d]	EA	E	SA	S	EA	E	SA	S
R^2	0.22	0.25	0.24	0.26	0.20	0.24	0.23	0.26	0.27	0.28	0.28	0.28
RMSEP (g kg^{-1})	24.4	23.8	24.0	23.5	24.6	23.9	24.1	23.6	23.4	23.3	23.2	23.2

[a] EA = exponential model with anisotropy

[b] E = exponential model

[c] SA = spherical model with anisotropy

[d] S = spherical model

2.6. Conclusions

VNIRS-PLSR with raw reflectance spectra successfully predicted SOC of Andisols within a micro-watershed of Costa Rica using a minimum number of soil analyses, with R^2 = 0.82, RMSEP = 11.7 g kg^{-1}, and RPD = 2.10. The method therefore provides 1) a rapid and cost effective analysis of SOC in complex agroforestry systems, 2) increased sample numbers to further analyze the relationships between SOC and landscape attributes as well as beneficial effects of agroforestry, and 3) more detailed understanding in the spatial structure of SOC within a micro-watershed.

For this study site, the significant correlations between SOC and landscape attributes were only apparent when the micro-watershed was stratified using SLU, presumably due to the differences in SOC concentrations between the SLU generated by soil water flow and sedimentation. Elevation had a significant negative correlation to SOC in upperslope and valley, which further suggests the effects of downward flow of water and redistribution of SOC-rich topsoil to lower elevations. The spatial structure of SOC at the study site indicated moderate spatial dependency; however, predicting SOC of unvisited sites with OK and COK methods was not sufficiently precise due to high nugget effects.

In summary, the SOC distribution in this micro-watershed is influenced by highly localized variability in organic matter input by coffee and shade trees, organic matter decomposition dynamics, volcanic ash content and allophane levels, and past and present agricultural management. This complicated spatial assessment of SOC stocks for the area. Using regression-based landscape models provided limited results, in part because the SOC distribution were not uniform and varied by landscape unit. Geostatistical techniques were similarly affected by high

local variability (nugget variance). The most promising results were achieved by VNIRS, which showed high predictability for SOC. It allows for better SOC stock estimation through high sample numbers.

Although the "paint by numbers" for SOC stock assessment approach may work well for cases with predictable spatial patterns, it appears that high spatial sampling intensities, inexpensive SOC analysis with VNIRS, and "measure and multiply" extrapolation is most useful for this complex agroforestry watershed.

2.7. Acknowledgements

This work was in part supported through Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD, France), the European project CAFNET (EuropAid/121998/C/G), USDA-NIFA Special Grant on Computational Agriculture, and the Cafetalera Aquiares farm (http://www.cafeaquiares.com). CATIE (Centro Agronómico Tropical de Investigación y Enseñanza) also provided laboratory facilities and collaborations. Dr. Dean Hively and the USDA-ARS laboratory kindly offered to analyze soil carbon and nitrogen. Samples were derived from sites established and maintained through EU Cafnet project, and the efforts of Dr. Olivier Roupsard, Dr. Federico Gómez-Delgado, Alexis Perez, Alvaro Barquero, Alejandora Barquero, and many field staff members and kind contributions from Aquiares farm.

REFERENCES

Arnold, S.L., J.W. Doran, J. Schepers, B. Wienhold, D. Ginting, B. Amos, and S. Gomes. 2005. Portable Probes to Measure Electrical Conductivity and Soil Quality in the Field. Communications in Soil Science and Plant Analysis 36:2271 - 2287.

Aspinall, R.J. 2002. Use of logistic regression for validation of maps of the spatial distribution of vegetation species derived from high resolution hyperspectral remotely sensed data. Ecological Modelling 157:301-312.

Awiti, A.O., M.G. Walsh, K.D. Shepherd, and J. Kinyamario. 2008. Soil condition classification using infrared spectroscopy: A proposition for assessment of soil condition along a tropical forest-cropland chronosequence. Geoderma 143:73-84.

Ball, D.F. 1964. Loss-on-Ignition as Estimate of Organic Matter + Organic Carbon in Non-Calcareous Soils. Journal of Soil Science 15:84-92.

Bastida, F., A. Zsolnay, T. Hernandez, and C. Garcia. 2008. Past, present and future of soil quality indices: A biological perspective. Geoderma 147:159-171.

Beer, J., R. Muschler, D. Kass, and E. Somarriba. 1998. Shade management in coffee and cacao plantations. Agroforestry Systems 38:139-164.

Ben-Dor, E., and A. Banin. 1995. Near-Infrared Analysis as a Rapid Method to Simultaneously Evaluate Several Soil Properties. Soil Science Society of America Journal 59:364-372.

Ben-Dor, E., J.R. Irons, and G.F. Epema. 1999. Soil reflectance, *In* N. Rencz, ed. Remote Sensing for the Earth Sciences: Mannual of Remote Sensing, Vol. 3. John Wiley & Sons, New York.

Beven, K. 1997. Topmodel: A Critique. Hydrological Processes 11:1069-1085.

Bilgili, A.V., F. Akbas, and H.M. van Es. 2011. Combined use of hyperspectral VNIR reflectance spectroscopy and kriging to predict soil variables spatially. Precision Agriculture 12:395-420.

Bilgili, A.V., H.M. van Es, F. Akbas, A. Durak, and W.D. Hively. 2010. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. Journal of Arid Environments 74:229-238.

Bishop, N.J., and A.B. McBratney. 2001. A comparison of prediction methods for the creation of field-extent soil property maps. Geoderma 103:149-160.

Bolstad, P.V., and J.M. Vose. 2001. The effects of terrain position and elevation on soil C in southern Appalanchians, *In* R. e. a. Lal, ed. Assessment methods for soil carbon. Lewis Publishers, Boca Raton, FL.

Bowers, S.A., and R.J. Hanks. 1965. Reflection of Radiant Energy From Soils. Soil Science 100:130-138.

Brunet, D., B.G. Barthes, J.L. Chotte, and C. Feller. 2007. Determination of carbon and nitrogen contents in Alfisols, Oxisols and Ultisols from Africa and Brazil using NIRS analysis: Effects of sample grinding and set heterogeneity. Geoderma 139:106-117.

Burgess, T.M., and R. Webster. 1980. Optimal Interpolation and Isarithmic Mapping of Soil Properties .1. The Semi-Variogram and Punctual Kriging. Journal of Soil Science 31:315-331.

Cambardella, C.A., A.T. Moorman, J.M. Novak, T.B. Parkin, D.L. Karlen, R.F. Turco, and A.E. Konopka. 1994. Field-scale heterogenity of soil properties in central Iowa soils. Soil Science Society of America Journal 58:192-194.

CENIGA. 1998. Hojas Topográficas Escala 1:25000. Proyect TERRA.

Chang, C.W., D.A. Laird, M.J. Mausbach, and C.R. Hurburgh. 2001. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. Soil Science Society of America Journal 65:480-490.

Chaplot, V., M. Bernoux, C. Walter, P. Curmi, and U. Herpin. 2001. Soil carbon storage prediction in temperate hydromorphic soils using a morphologic index and digital elevation model. Soil Science 166:48-60.

Chesworth, W. 2008. Encyclopedia of Soil Science Springer, Dordrecht, The Netherlands.

Chikhaoui, M., F. Bonn, A.I. Bokoye, and A. Merzouk. 2005. A spectral index for land degradation mapping using ASTER data: Application to a semi-arid Mediterranean catchment. International Journal of Applied Earth Observation and Geoinformation 7:140-153.

Conant, R.T., G.R. Smith, and K. Paustian. 2003. Spatial variability of soil carbon in forested in cultivated sites: Implications for change detection. Journal of Environmental Quality 32:278-286.

Couteaux, M.M., L. Sarmiento, D. Herve, and D. Acevedo. 2005. Determination of water-soluble and total extractable polyphenolics in biomass, necromass and decomposing plant material using near-infrared reflectance spectroscopy (NIRS). Soil Biology & Biochemistry 37:795-799.

Cox, T.L., Harris, W.F., Asmus, B.S., & Edwards, N.T. (1978). The role of fine roots in biogeochemical cycles in eastern deciduous forest. Pedobiologia, 18, 264-271

Cozzolino, D., and A. Morón. 2003. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. Journal of Agricultural Science 140.

Creaser, C.S., and A.M.C. Davies. 1988. Chemometrics and data analysis Analytical applications of spectroscopy. Royal Society of Chemistry, London.

Dalal, R.C., and R.J. Henry. 1986. Simultaneous Determination of Moisture, Organic-Carbon, and Total Nitrogen by near-Infrared Reflectance Spectrophotometry. Soil Science Society of America Journal 50:120-123.

Dardanne, P., G. Sinnaeve, and V. Baeten. 2000. Multivariate calibration and chemometrics for near infrared spectroscopy: which method? Journal of near Infrared Spectroscopy 8:229-237.

Dexter, A.R. 2004. Soil physical quality - Part 1. Theory, effects of soil texture, density, and organic matter, and effects on root growth. Geoderma 120:201-214.

Dilts, T. 2010. Topography Tools for ArcGIS [Online] http://arcscripts.esri.com/details.asp?dbid=15996.

Doran, J.W., and T.B. Parkin. 1994. Defining and assessing soil quality, *In* J. W. Doran, et al., eds. Defining Soil Quality for a Sustainable Environment. Soil Science Society of America, Madison, WI.

Ersahin, S. 2003. Comparing ordinary kriging and cokriging to estimate infiltration rate. Soil Science 67:1848-1855.

Eswaran, H., E. van den Berg, P. Reich, and J. Kimble. 1995. Global soil carbon resources, p. 27-43, *In* R. e. a. Lal, ed. Soils and global change. Adv. Soil Sci. CRC Press, Boca Raton, FL.

Florinsky, I.V., R.G. Eilers, G.R. Manning, and L.G. Fuller. 2002. Prediction of soil properties by digital terrain modelling. Environmental Modelling & Software 17:295-311.

Fournier, L.A. 1988. El cultivo del café (Coffea arabica L.) al sol o a la sombra: un enfoque agronómico y eco fisiológico. Agronomía Costarricense 12:131-146.

Fystro, G. 2002. The prediction of C and N content and their potential mineralisation in heterogeneous soil samples using Vis-NIR spectroscopy and comparative methods. Plant and Soil 246:139-149.

Garten, J.C.T., I.W.M. Post, P.J. Hanson, and L.W. Cooper. 1999. Forest soil carbon inventories and dynamics along an elevation gradient in the southern Appalachian Mountains. Biogeochemistry 45:115-145.

Gessler, P.E., O.A. Chadwick, F. Charman, L. Althouse, and K. Holmes. 2000. Modeling Soil-Landscape and Ecosystem Properties Using Terrain Attributes. Soil Science Society of America Journal 64:2046-2056.

Gómez-Delgado, F. 2010. Hydrological, ecophysiological and sediment processes in a coffee agroforestry basin: combining experimental and modelling methods to assess hydrological environmental services., Supagro, Montpellier.

Gómez-Delgado, F., O. Roupsard, R. Moussa, G. le Maire, S. Taugourdeau, J.M. Bonnefond, A. Pérez, M. van Oijen, P. Vaast, B. Rapidel, M. Voltz, P. Imbach, and J.M. Harmand. 2011. Modelling the hydrological behaviour of a coffee agroforestry basin in Costa Rica. Hydrology and Earth System Sciences 15:369-392.

Goovaerts, P. 1997. Geostatistics for natural resources evaluation Oxford University Press, New York.

Gower, S.T., C.J. Kucharik, and J.M. Norman. 1999. Direct and Indirect Estimation of Leaf Area Index, fAPAR, and Net Primary Production of Terrestrial Ecosystems. Remote Sensing of Environment 70:29-51.

Grier, C.C., Vogt, K.A., Keyes, M.R., & Edmonds, R.L. (1981). Biomass distribution and above- and below-ground production in yound and mature *Abies amabilis* zone ecosystems of the Washington Cascades. Canadian Journal of Forest Research, 11, 155-167

Gugino, B.K., O.J. Idowu, R.R. Schindelbeck, H.M. van Es, D.W. Wolfe, B.N. Moebius-Clune, J.E. Thies, and G.S. Abawi. 2009. Cornell Soil Health Assessment Training Manual, Edition 2.0 Cornell University, Geneva, NY.

Hart, J.R., C. Golumbic, and K.H. Norris. 1962. Determination of moisture content of seeds by near-infrared spectrophotometry of their methanol extracts. Cereal Chemistry 39:94-&.

Heise, H.M., L. Kupper, and L.N. Butvia. 2002. Bio-analytical applications of mid-infrared spectroscopy using silver halide fiber-optic probes. Spectrochimica Acta Part B-Atomic Spectroscopy 57:1649-1663.

Hengl, T., G.B.M. Heuvelink, and A. Stein. 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120:75-93.

Herrick, J., and W. Whitford. 1995. Assessing the quality of rangeland soils: challenges and opportunities. Journal of Soil and Water Conservation 50:237-242.

Hively, W.D., G. McCarty, J.B. Reeves Ⅲ, M.W. Lang, R.A. Osterling, and S.R. Delwiche. 2011. Use of airborne hyperspectral imagery to map soil properties in tilled agricultural fields. Applied and Environmental Soil Science.

Holst, H. 1992. Comparison of different calibration methods suited for calibration problems with many variables. Applied Spectroscopy 46:1780-1784.

Hoyos, N., and N.B. Comerford. 2005. Land use and landscape effects on aggregate stability and total carbon of Andisols from the Colombian Andes. Geoderma 129:268-278.

Hunt, G.R. 1980. Spectroscopic properties of rock and minerals, p. 295, *In* C. R. Stewart, ed. Handbook of physical properties of rocks. CEC Press Inc, Florida.

Idowu, O.J., H.M. van Es, G.S. Abawi, D.W. Wolfe, J.I. Ball, B.K. Gugino, B.N. Moebius, R.R. Schindelbeck, and A.V. Bilgili. 2008. Farmer-oriented assessment of soil quality using field, laboratory, and VNIR spectroscopy methods. Plant and Soil 307:243-253.

Isaaks, E.H., and R.M. Srivastava. 1989. An introduction to applied geostatistics Oxford University Press, New York.

Islam, K.R., and R.R. Weil. 2000. Land use effects on soil quality in a tropical forest ecosystem of Bangladesh. Agriculture Ecosystems & Environment 79:9-16.

Jenness, J. 2006. Topographic Position Index (tpi_jen.avx) extension for ArcView 3.x, v. 1.2. Jenness Enterprises.

Jian-Bing, W., X. Du-Ning, Z. Xing-Yi, L. Xiu-Zhen, and L. Xiao-Yu. 2006. Spatial variability of soil organic carbon in relation to environmental factors of a typical small watershed in the black soil region, northeast China. Environ Monit Assess 121:597-613.

Johnston, C.T., W.M. Davies, C. Erickson, J.J. Delfino, and W.T. Cooper. 1994. Characterization of humic substances using Fourier transform infrared spectroscopy, *In* N. Senesi and T. Miano, eds. Humic substances in the global environment and implications on human health. Elsevier Science, Amsterdam.

Journel, A.G., and Ch.J. Huijbregts. 1978. Mining geostatistics Academic Press, New York.

Karlen, D.L., N.C. Wollenhaupt, D.C. Erbach, E.C. Berry, J.B. Swan, N.S. Eash, and J.L. Jordahl. 1994. Crop residue effects on soil quality following 10 years of no-till corn. Soil and Tilage Research 31:149-167.

Kettler, T.A., J.W. Doran, and T.L. Gilbert. 2001. Simplified method for soil particle-size determination to accompany soil-quality analyses. Soil Science Society of America Journal 65:849-852.

Kimetu, J.M., J. Lehmann, S.O. Ngoze, D.N. Mugendi, J.M. Kinyangi, S. Riha, L. Verchot, J.W. Recha, and A.N. Pell. 2008. Reversibility of soil productivity decline with organic matter of differing quality along a degradation gradient. Ecosystems 11:726-739.

Kinoshita, R., B.N. Moebius-Clune, H.M. van Es, W.D. Hively, and A.V. Bilgili. 2011: in review. Strategies for Soil Quality Assessment Using VNIR Hyperspectral Spectroscopy in a Western Kenya Chronosequence. Soil Science Society of America Journal.

Kinyangi, J. 2008. Soil degradation, thresholds and dynamics of long-term cultivation: from landscape biogeochemistry to nanoscale biogeocomplexity, Cornell University, Ithaca, NY.

Kooistra, L., R. Wehrens, R.S.E.W. Leuven, and L.M.C. Buydens. 2001. Possibilities of visible-near-infrared spectroscopy for assessment of soil contamination in river floodplains. Analytica Chimica Acta 446:97-105.

Kooistra, L., J. Wanders, G.F. Epema, R.S.E.W. Leuven, R. Wehrens, and L.M.C. Buydens. 2003. The potential of field spectroscopy for the assessment of sedient properties in river floodplains. Analytica Chimica Acta 484:189-200.

Kravchenko, A.N., and G.P. Robertson. 2007. Can topographical and yield data substantially improve total soil carbon mapping by regression kriging? Agronomy Journal 99:12-17.

Krull, E.S., E.A. Bestland, and W.P. Gates. 2002. Soil organic matter decomposition and turnover in a tropical Ultisol: Evidence from delta C-13, delta N-15 and geochemistry. Radiocarbon 44:93-112.

Kusumo, B.H., M.J. Hedley, M.P. Tuohy, C.B. Hedley, and G.C. Arnold. 2010. Predicting soil carbon and nitrogen concentrations and pasture root densities from proximally sensed soil spectral reference, p. 177-190, *In* R. A. Viscarra Rossel, et al., eds. Proximal soil sensing. Springer, Dordrecht ; London.

Lal, R. 2006. Enhancing crop yields in the developing countries through restoration of the soil organic carbon pool in agricultural lands. Land Degradation & Development 17:197-209.

Lamsal, S. 2009. Visible Near-Infrared Reflectance Spectrocopy for Geospatial Mapping of Soil Organic Matter Soil Science 174:35-44.

le Maire, G., C. Francois, K. Soudani, H. Davi, V. Le Dantec, B. Saugier, and E. Dufrene. 2006. Forest leaf area index determination: A multiyear satellite-independent method based on

within-stand normalized difference vegetation index spatial variability. Journal of Geophysical Research-Biogeosciences 111.

Lung, T., and G. Schaab. 2006. Assessing fragmentation and disturbance of west Kenyan rainforests by means of remotely sensed time series data and landscape metrics. African Journal of Ecology 44:491-506.

Magdoff, F.R., and H.M. van Es. 2009. Building Soils for Better Crops Sustainable Agric. Research and Extension, College Park, MD.

Matheron, G. 1978. Recherche de simplification dans un problème de cokrigeage Centre de Geostatistique, Fontainbleau.

McCarty, G.W., and J.B. Reeves. 2006. Comparison of Near Infrared and Mid Infrared Diffuse Reflectance Spectroscopy for Field-Scale Measurement of Soil Fertility Parameters. Soil Science 171:94-102 10.1097/01.ss.0000187377.84391.54.

McKenzie, N.J., H.P. Cresswell, P.J. Ryan, and M. Grundy. 2000. Contemporary land resource survey requires improvements in direct soil measurement. Communications in Soil Science and Plant Analysis 31:1553-1569.

McLachlan. 1992. Discriminant Analysis and Statistical Pattern Recognition John Wiley & Sons, Hoboken, NJ.

Mehlich, A. 1984. Mehlich-3 soil test extractant - a modification of mehlich-2 extractant. Communications in Soil Science and Plant Analysis 15:1409-1416.

Millenium Ecosystem Assessment. 2005. Ecosystem and Wellbeing: Synthesis. Island Press, Washington DC.

Moebius, B.N., H.M. van Es, R.R. Schindelbeck, O.J. Idowu, J.E. Thies, and D.J. Clune. 2007. Evaluation of Laboratory-Measured Soil Properties as Indicators of Soil Physical Quality. Soil Science 172:895-912.

Moebius-Clune, B.N. 2010. Applications of integrative soil quality assessment in research, extention, and eduaction, Cornell University, Ithaca, NY.

Moebius-Clune, B.N., H.M. van Es, O.J. Idowu, R.R. Schindelbeck, J.M. Kimetu, S. Ngoze, J. Lehmann, and J. Kinyangi. 2011. Long-term soil quality degradation along a cultivation chronosequence in Western Kenya. Agriculture Ecosystems & Environment.

Moore, I.D., R.B. Grayson, and A.R. Ladson. 1991. Digital Terrain Modeling - a Review of Hydrological, Geomorphological, and Biological Applications. Hydrological Processes 5:3-30.

Moore, I.D., P.E. Gessler, G.A. Nielsen, and G.A. Peterson. 1993. Soil Attribute Prediction Using Terrain Analysis. Soil Science Society of America Journal 57:443-452.

Mora-Chinchilla, R. 2000. Geomorfología de la cuenca del Río Turrialba. Universidad de Costa Rica, San José, Costa Rica.

Morgan, C.L.S., T.H. Waiser, D.J. Brown, and C.T. Hallmark. 2009. Simulated in situ characterization of soil organic and inorganic carbon with visible near-infrared diffuse reflectance spectroscopy. Geoderma 151:249-256.

Morón, A., and D. Cozzolino. 2002. Application of near infrared reflectance spectroscopy for the analysis of organic C, total N and pH in soils of Uruguay. Journal of near Infrared Spectroscopy 10:215-221.

Morón, A., and D. Cozzolino. 2004. Determination of potentially mineralizable nitrogen and nitrogen in particulate organic matter fractions in soil by visible and near-infrared reflectance spectroscopy. Journal of Agricultural Science 142:335-343.

Mtambanengwe, F., P. Mapfumo, and B. Vanlauwe. 2006. Comparative short-term effects of different quality organic resources on maize productivity under two different environments in Zimbabwe. Nutrient Cycling in Agroecosystems 76:271-284.

Muschler, R.G. 2001. Shade improves coffee quality in a sub-optimal coffee-zone of Costa RIca. Agroforestry Systems 85:131-139.

Naes, T., T. Isaksson, and B. Kowalski. 1990. Locally weighted regression and scatter correction for nea-infrared reflectance data. Analytical Chemistry 62:664-673.

Nanzyo, M., R. Dahlgren, and S. Shoji. 1993. Chemical characteristics of volcanic ash soils, *In* S. Shoji, et al., eds. Volcanic Ash Soils - Genesis, Properties and Utilization. Elsevier Science Publishers B.V., Amsterdam.

Ngoze, S., S. Riha, J. Lehmann, L. Verchot, J. Kinyangi, D. Mbugua, and A. Pell. 2008. Nutrient constraints to tropical agroecosystem productivity in long-term degrading soils. Global Change Biology 14:2810-2822.

Ogden, C.B., H.M. vanEs, and R.R. Schindelbeck. 1997. Miniature rain simulator for field measurement of soil infiltration. Soil Science Society of America Journal 61:1041-1043.

Payán, F., D.L. Jones, J. Beer, and J.M. Harmand. 2009. Soil characteristics below Erythrina poeppigiana in organic and conventional Costa Rican coffee plantations. Agroforestry Systems 76:81-93.

Peel, M.C., B.L. Finlayson, and T.A. McMahon. 2007. Updated world map of the Koppen-Geiger climate classification. Hydrology and Earth System Sciences 11:1633-644.

Peltre, C., L. Thuriès, B. Barthès, D. Brunet, T. Morvan, B. Nicolardot, V. Parnaudeau, and S. Houot. 2011. Near infrared reflectance spectroscopy: A tool to characterize the composition of different types of exogenous organic matter and their behaviour in soil. Soil Biology & Biochemistry 43:197-205.

Pendergrast, M. 2009. Coffee: Second to Oil? Tea & Coffee Trade Journal 181:38-41.

Post, W.M., R.C. Izaurralde, L.K. Mann, and N. Bliss. 2001. Monitoring and verifying changes of organic carbon in soil. Climatic Change 51:73-99.

Reeves, J., G. McCarty, and T. Mimmo. 2002. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. Environmental Pollution 116:S277-S284.

Reeves, J.B., and J.S. Van Kessel. 1999. Investigations into near-infrared analysis as an alternative to traditional procedures in manure N and C mineralization studies. Journal of near Infrared Spectroscopy 7:195-212.

Reeves, J.B., R.F. Follett, G.W. McCarty, and J.M. Kimble. 2006. Can near or mid-infrared diffuse reflectance spectroscopy be used to determine soil carbon pools? Communications in Soil Science and Plant Analysis 37:2307-2325.

Reeves, J.B., G.W. Mccarty, F. Calderon, and W.D. Hively. 2011: in review. Advances in spectroscopic methods for quantifying soil carbon, *In* M. A. Liebig, et al., eds. Managing Agricultural Greenhouse Gases. Elsevier Press, Amsterdam.

Rhodes, E.R., P.Y. Kamara, and P.M. Sutton. 1981. Walkley-Black digestion efficiency and relationship to loss on ignition for selected Sierra Leone soils. Soil Science Society of America Journal 45:1132-1135.

Rouse, J.W., R.H. Haas, J.A. Schell, and D.W. Deering. 1973. Monitoring vegetation systems in the Great Plains with ERTS, p. 309-317 Third ERTS Symposium, Vol. 1. NASA.

Ruess, R.W., Van Cleve, K., Yarie, J., & Uiereck, L.A. (1996). Contributions of fine root production and turnover to the carbon and nitrogen cycling in taiga forests of the Alaskan interior. Canadian Journal of Forest Research, 26, 1326-1336

Ruffin, C., and R.L. King. 1999. The analysis of hyperspectral data using Savitzky-Golay filtering - Theoretical basis (part 1) Geoscience and Remote Sensing Symposium, 1999, Hamburg, Germany.

Russell, C.A. 2003. Sample preparation and prediction of soil organic matter properties by near infra-red reflectance spectroscopy. Communications in Soil Science and Plant Analysis 34:1557-1572.

Saetre, P. 1999. Spatial patterns of ground vegetation, soil microbial biomass and activity in a mixed spruce-birch stand. Ecography 22:183-192.

Salgó, A., J. Nagy, J. Tarnóy, P. Marth, O. Pálmai, and G. Szabó-Kele. 1998. Characterisation of soils by the near infrared technique. Journal of near Infrared Spectroscopy 6:199.

Sanchez, P.A. 2002. Ecology–soil fertility and hunger in Africa. Science 295:2019-2020.

Savitzky, A., and M.J.E. Golay. 1964. Smoothing and differentiation of data by simplified least square procedure. Analytical Chemistry 36:1627-1639.

Schimel, D.S., and C.S. Potter. 1995. Process modelling and spatial extrapolation, p. 358-383, *In* P. A. Matson and R. C. Harriss, eds. Biogenic trace gases: Measuring emissions from soil and water. Blackwell Science Ltd., Cambridge, MA.

Shepherd, K.D., and M.G. Walsh. 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Science Society of America Journal 66:988-998.

Solomon, D., J. Lehmann, J. Kinyangi, W. Amelung, I. Lobe, A. Pell, S. Riha, S. Ngoze, L. Verchot, D. Mbugua, J. Skjemstad, and T. Schafer. 2007. Long-term impacts of anthropogenic perturbations on dynamics and speciation of organic carbon in tropical forest and subtropical grassland ecosystems. Global Change Biology 13:511-530.

Soussana, J.F., P. Louiseau, N. Vuichard, E. Ceschia, J. Balesdent, T. Chevallier, and D. Arrouays. 2004. Carbon cycling and sequestration opportunities in temperate grasslands. Soil Use and Management 20:219-230.

Stevens, A., B. van Wesemael, H. Bartholomeus, D. Rosillon, B. Tychon, and E. Ben-Dor. 2008. Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils Geoderma 144:395-404.

Stoner, E.R., and M.F. Baumgardner. 1981. Characteristic variations in reflectance of surface soils. Soil Science Society of America Journal 45:1161-1165.

Takata, Y., S. Funakawa, K. Akshalov, N. Ishida, and T. Kosaki. 2007. Spatial prediction of soil organic matter in northern Kazakhstan based on topographic and vegetation information. Soil Science and Plant Nutrition 53:289-299.

Tate, K.R., and D.J. Ross. 1997. Elevated CO2 and moisture effects on soil carbon storage and cycling in temperate grasslands. Global Change Biology 3:225-235.

Taugourdeau, S., G. le Maire, O. Roupsard, J. Avelino, F. Gómez-Delgado, J.R. Jones, C. Marsden, A. Robelo, E. Alpizar, A. Barquero, B. Rapidel, P. Vaast, and J.M. Harmand. 2010. Scaling-up LAI in coffee agroforestry systems in Costa Rica ASIC 2010, Bali, Indonesia

Terra, J.A., J.N. Shaw, D.W. Reeves, R.L. Raper, E. van Santen, and P.L. Mask. 2004. Soil Carbon Relationships With Terrain Attributes, Electrical Conductivity, and A Soil Survey in A Coastal Plain Landscape. Soil Science 169:819-831.

The World Bank. 2010. Kenya: Agricultural Carbon Project [Online]. Available by The World Bank http://wbcarbonfinance.org/Router.cfm?Page=BioCF&FID=9708&ItemID=9708&ft=Projects&ProjID=58099 (verified 06/23/2011).

Thomasson, J.A., R. Sui, M.S. Cox, and A. Al-Rajehy. 2001. Soil reflectance sensing for determining soil properties in precision agriculture. Transactions of the Asae 44:1445-1453.

Thompson, J.A., and R.K. Kolka. 2005. Soil carbon storage estimation in a forest watershed using quantitative soil-landscape modeling. Soil Science Society of America Journal 69:1086-1093.

Topp, G.C., Y.T. Galganov, B.C. Ball, and M.R. Carter. 1993. Soil water desorption curves, *In* M. R. Carter, ed. Soil sampling and methods of analysis. Canadian Society of Soil Science, Lewis Publishers, Boca Raton, FL.

Udelhoven, T., C. Emmerling, and T. Jarmer. 2003. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial-least square regression: a feasibility study. Plant and Soil 251:319-329.

Vaast, P., R.F. van Kanten, P. Siles, J. Angrand, and A. Aguilar. 2007. Biophysical interactions between timber trees and Arabica coffee in suboptimal conditions of Central America., p. 136-148, *In* S. Jose and A. Gordon, eds. Toward Agroforestry Design: An Ecological Approach. Springer.

Vagen, T.G., K.D. Shepherd, and M.G. Walsh. 2006. Sensing landscape level change in soil fertility following deforestation and conversion in the highlands of Madagascar using Vis-NIR spectroscopy. Geoderma 133:281-294.

Variangis, P., P. Siegel, D. Giovannuci, and B. Lewin. 2003. Dealing with coffee crisis in Central America: Impacts and strategies. The World Bank, Washington DC.

Vasques, G.M., S. Grunwald, and J.O. Sickman. 2009. Modeling of soil organic carbon fractions using visible-near-infrared spectroscopy. Soil Science Society of America Journal 73:176-184.

Viera, A.J., and J.M. Garrett. 2005. Understanding interobserver agreement: the kappa statistic. Family Medicine 37:360-363.

Viscarra Rossel, R.A., and A.B. McBratney. 1998a. Laboratory evaluation of a proximal sensing technique for simultaneous measurement of soil clay and water content. Geoderma 85:19-39.

Viscarra Rossel, R.A., and A.B. McBratney. 1998b. Soil chemical analytical accuracy and costs: implications from precision agriculture. Australian Journal of Experimental Agriculture 38:765-775.

Viscarra Rossel, R.A., D.J.J. Walvoort, A.B. McBratney, L.J. Janik, and J.O. Skjemstad. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131:59-75.

Webster, R., and M.A. Oliver. 2001. Geostatistics for environmental scientists John Wiley & Sons, Chichester, England ; New York.

Weil, R.R., K.R. Islam, M.A. Stine, J.B. Gruver, and S.E. Samson-Liebig. 2003. Estimating active carbon for soil quality assessment: A simplified method for laboratory and field use. American Journal of Alternative Agriculture 18:3-17.

Williams, P.C. 1975. Applications of near-infrared reflectance spectroscopy to analysis of cereal grains and oilseeds. Cereal Chemistry 52:561-576.

Williams, P.C., and D.C. Sobering. 1993. Comparison of commercial near infra-red transmittance and reflectance instruments for analysis of whole grains and seeds. Journal of near Infrared Spectroscopy 1:25-33.

Wongpokhom, N., I. Kheoruenromne, A. Suddhiprakarn, and R.J. Gilkes. 2008. Micromorphological properties of salt affected soils in Northeast Thailand. Geoderma 144:158-170.